

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 13:56:39 ; Search time 6938.09 Seconds
(without alignments)
8578.019 Million cell updates/sec

Title: US-09-776-865-3
Perfect score: 2844
Sequence: 1 cccggggggcggggggtctg.....ataaaaaaaaaaaaaaaaaa 2844

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgc_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| ----- | | | | | |

| | | | | | | |
|----|--------|-------|--------|----|-----------|---------------------|
| 1 | 2844 | 100.0 | 2844 | 4 | AF244578 | AF244578 Ovis arie |
| 2 | 2844 | 100.0 | 2844 | 6 | AX207626 | AX207626 Sequence |
| 3 | 1572.4 | 55.3 | 2512 | 6 | AX138494 | AX138494 Sequence |
| 4 | 1572.4 | 55.3 | 2512 | 9 | HSAB37747 | AJ387747 Homo sapi |
| 5 | 1570.8 | 55.2 | 2930 | 6 | AX207624 | AX207624 Sequence |
| 6 | 1570.8 | 55.2 | 2930 | 6 | AF244577 | AF244577 Homo sapi |
| 7 | 1570.8 | 55.2 | 3329 | 9 | BC020961 | BC020961 Homo sapi |
| 8 | 792.2 | 27.9 | 1587 | 9 | AK025880 | AK025880 Homo sapi |
| 9 | 602.8 | 21.2 | 2006 | 9 | AK026921 | AK026921 Homo sapi |
| 10 | 480.2 | 16.9 | 56641 | 2 | AL138833 | AL138833 Homo sapi |
| 11 | 480.2 | 16.9 | 113202 | 9 | HS397H23 | AL121972 Human DNA |
| 12 | 480.2 | 16.9 | 149597 | 2 | AC034471 | AC034471 Homo sapi |
| 13 | 480.2 | 16.9 | 157749 | 2 | AC025535 | AC025535 Homo sapi |
| 14 | 314.2 | 11.0 | 375 | 6 | AX118967 | AX118967 Sequence |
| 15 | 269.6 | 9.5 | 3616 | 9 | AK024903 | AK024903 Homo sapi |
| 16 | 268 | 9.4 | 2626 | 2 | AB020527 | AB020527 Homo sapi |
| 17 | 261.4 | 9.2 | 179129 | 2 | AC097023 | AC097023 Rattus no |
| 18 | 230.2 | 8.1 | 2004 | 10 | BC013445 | BC013445 Mus muscu |
| 19 | 228.6 | 8.0 | 1885 | 10 | MMNPR1CT | XJ7241 M.musculus |
| 20 | 226.6 | 8.0 | 1841 | 3 | AY060776 | AY060776 Drosophill |
| 21 | 225.4 | 7.9 | 3946 | 9 | AB032435 | AB032435 Homo sapi |
| 22 | 223.2 | 7.8 | 1700 | 10 | RNU28504 | U28504 Rattus norv |
| 23 | 219 | 7.7 | 1865 | 10 | BC018252 | BC018252 Mus muscu |
| 24 | 217.4 | 7.6 | 3982 | 10 | AF324864 | AF324864 Mus muscu |
| 25 | 217.4 | 7.6 | 3982 | 10 | AF271235 | AF271235 Rattus no |
| 26 | 210.2 | 7.4 | 1549 | 6 | AX336195 | AX336195 Sequence |
| 27 | 210.2 | 7.4 | 1549 | 6 | AX336653 | AX336653 Sequence |
| 28 | 210.2 | 7.4 | 1549 | 9 | HSNAP11 | XJ7355 H.sapiens m |
| 29 | 209.8 | 7.4 | 1874 | 10 | BC018306 | BC018306 Mus muscu |
| 30 | 205.4 | 7.2 | 1794 | 9 | HUAP14 | D28532 Human mRNA |
| 31 | 193.2 | 6.8 | 2024 | 10 | RNU07609 | U07609 Rattus norv |
| 32 | 192.6 | 6.8 | 2748 | 3 | AY069501 | AY069501 Drosophill |
| 33 | 190.6 | 6.7 | 1710 | 3 | AY075277 | AY075277 Drosophill |
| 34 | 190.6 | 6.7 | 34229 | 2 | AC014600 | AC014600 Drosophill |
| 35 | 190.6 | 6.7 | 267488 | 3 | AE003640 | AE003640 Drosophill |
| 36 | 190.6 | 6.7 | 303885 | 9 | DROSADH01 | AE003407 Drosophill |
| 37 | 186.8 | 6.6 | 2366 | 3 | AB032436 | AB032436 Homo sapi |
| 38 | 185.2 | 6.5 | 2716 | 6 | I40028 | I40028 Sequence 1 |
| 39 | 185.2 | 6.5 | 2716 | 6 | I40029 | I40029 Sequence 1 |
| 40 | 185.2 | 6.5 | 2716 | 6 | I40220 | I40220 Sequence 1 |
| 41 | 185.2 | 6.5 | 2716 | 6 | I40221 | I40221 Sequence 1 |
| 42 | 185.2 | 6.5 | 2716 | 6 | I73259 | I73259 Sequence 1 |
| 43 | 185.2 | 6.5 | 2716 | 6 | I73260 | I73260 Sequence 1 |
| 44 | 183.6 | 6.5 | 2178 | 3 | AF095787 | AF095787 Caenorhab |
| 45 | 180.4 | 6.3 | 1855 | 4 | RABRCTX | M76466 Rabbit rena |

ALIGNMENTS

RESULT 1
AF244578
LOCUS AF244578 2844 bp mRNA linear MAM 06-AUG-2000
DEFINITION Ovis aries membrane glycoprotein SP55 (sp55) mRNA, complete cds.
ACCESSION AF244578
VERSION AF244578.1 GI:9719375
KEYWORDS
SOURCE
ORGANISM
Ovis aries
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Caprinae; Ovis.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 2844)
Fu,C., Bardhan,S., Cetateanu,N.D., Lloyd,S.R., Yan,H.-P.,
Carter,C.E., Shl,E., Venkov,C., Yakes,M.F., Page,D.L. and H.C.G.
Identification of a novel membrane protein from mammalian cells
that interact with the anti-pathoangogenic compound CM101
Unpublished
of Medicine, 23rd@Pierce, Nashville, TN 37232-0146, USA

FEATURES
source Location/Qualifiers
1. 2844
/organism="Ovis aries"
/db_xref="taxon:9940"
/cell_type="endothelial cell"
/tissue_type="lung"
1. 2844
/gene="sp55"
84. 1571
/gene="sp55"
/note="similar to Na+ inorganic phosphate transporter"
/codon_start=-1
/product="membrane glycoprotein SP55"
/protein_id="AF97770.1"
/db_xref="GI:9719376"
/translation="MKSPPSDLAAPSDGEEGRPTLLORAPRAPEAPVCCSARYNIAP
LSPEFPVLVSLFPLVNLVVALVDVDSNTAKDRRTSECEHSAPIKVLHNOTGKTR
WDAETOGWILGFFFGITITQIPGIVASHSGKLLGFEIFATLFTPLAADFG
VGAVALRALLEGSEVTPAMHAMSSVMPLEKRLKLSIYAGQLGVSLPLSG
VICVYNNMTYVEYFEGIVGIMFIMICLVDSPETPKETPEKEIISLKNLS
OKSVPMIPMLKSLPLMAIVVAHFSYNNTEFTLLTLPYKELYRNIENGFLSAP
YLGCWLCMLISGOADNLRARNFSTLWNRVRSILGIMGPAIEIVAAGTIGDYSIA
VARLTISTLIGRCSGSFSINHDIAVSAGILGITNPATIPGMIGPIIASLPIE
NTIGENOTVPCIAAINVFAIFFTLFAKEGVQNMALSDHGRN"

BASE COUNT 745 a 623 c 594 g 882 t
ORIGIN

Query Match 100.0%; Score 2844; DB 4; Length 2844;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2844; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccggggcggggggcttcggcggtcccgctgagctctcttcccgcgagagcagtttg 60
Db 1 CCCGGGGCGGGGGGCTTCGGCGGCTCCCGTGAGCTCTCTTCCCGGAGCAGGTTTG 60

QY 61 cgcgcgtagctccctgaagcaatcatgaagtcgccggtttcgacttagcccgagcag 120
Db 61 CGCGGTAGCTCCCGAAGGAGATCATGAAGTCCCGGTTTGGACTTGAAGCCGAGCAGC 120

QY 121 gcgagggagggctcgagcagcaaccgctctctgacgcgcccccgggcggaaccgctc 180
Db 121 GCGAGGAGGGCTCGAGCCGACACCGCTCTCTGACGCGCCCCCGGGGAGAACCGCTC 180

QY 181 cagtaatgctcgtcgtcgttacaacacttagcatlttgccttlttgggttcttcgtc 240
Db 181 CAGTATGCTGCTGCTGCTGTTACAACTTAGCATTTTGTCTTTTGGTTTCTTCTTTC 240

QY 241 tctattcaatcaggggtgaatctgagcgttgcaatgtgacatgtgtgatltcaacaa 300
Db 241 TCTATTCAATTAAGGGTGAATCTGAGCGTTGCACTAGTGAGCATGTGTGATTCAAACAA 300

QY 301 ctgcgaagaataagagctcctcctaagagtgtgagagcaatttcgtcccaataaagttc 360
Db 301 CTGCGAAGAATATAGACGTCTCTACAGAGTGTCAGAGCATTTCTGCTCCATTAATAAGTT 360

QY 361 ttcaacaacaaacgggtaaaaagtaaccggtggagtgcagaacactcaagagatgtatctcg 420
Db 361 TTCAACAACAACGGGTAAAAAGTAAACGGGTGAGATGCAGAAACTCAAGAGATGTTCG 420

QY 421 gatcttttctatgtcatcatcaacacaattctcgtgagagatagtgtgccagcaaa 480
Db 421 GATCTTTTCTATGTGCTATCATCATCACAAATTTCTGAGAGATATGTTGCCAGCAAA 480

QY 481 gtggggggaagcgtgtgctgagattcgggatctttgtlaagctatcttcaacctgttca 540
Db 481 GTGGGGGGAAGCTGTGCTAGGATTCGGGATCTTTGCTTACAGCATTCATACCTGTGTTCA 540

QY 541 ctccccctgcgtcagatcttcgaggtcgagccctgtgtgcaactcaaggagcactagaagggc 600
Db 541 CTCCCCCTGCCTGCAGATTTTGGAGTGGAGCCCTGTGTTGCACATCAGGCGCATAGAGGCG 600

QY 601 taggagaggggtgtcaacataaccagcaatgtccatgtgttctcatatgggtctccctc 660

Db 601 TAGAGAGGGGTGTCATATCCACAGCCATGCAATGCGATGTGCTTTCATGAGCTCCCTC 660

QY 661 ttgaaagaagcaacttcgtgatatctcatatgtagagcaacaacttggagcagtaattc 720
Db 661 TTGAAAGAAGCAACTTCGTGATATTTTCATATGAGAGCAACAACCTTGGGACATAGATT 720

QY 721 cttctcctcttcgtgagtaatttgcatacctaagaattggaactatgtcttctatctc 780
Db 721 CTCCTCTCTTTCGTGAGTAAATTTGCACTATATGAATTTGCACTTATGCTCTCATTTCT 780

QY 781 ttggcattgtggaatcaatcgtgttlaattlaagatcgtccttagttagtataccag 840
Db 781 TTGGCATTTGTGAATCATGTGCTTTATTTATGACTGCTGTTAGTATGATACACAG 840

QY 841 aaactcaagaacatactaccoccatbaaaaagagatattcttcttcaacttaaaaatc 900
Db 841 AAATCTCAAGACATATACCTCCATGAAAAAGAGATATTTCTTCAATTTAAAAAATC 900

QY 901 agctctctcaagaagtcagtgccgtggaatacctaagtcgaatcaactgcacacttg 960
Db 901 AGCTCTCTTCAAGAAATCATGTCGGTGATACCTATGCTGAATTCATCCTGCACTTGGG 960

QY 961 ctatgtcgttgcacatttctttaaaccctggacttctaactgtgtgaacttatgc 1020
Db 961 CTATTTGCTGTGCACATTTTCTTACAACTGAGACTTTTATCTTGTGACCTTATTTGC 1020

QY 1021 ctacttcatgaagaagtccttaagttcaattcaattcaagaagaagtttttcttcg 1080
Db 1021 CTACTTCAATGAAAGAGTCCATAGGTTCAATTTCAAGAGAAAGGATTTTATCTGACG 1080

QY 1081 tcccttattagtttgttgttlaatgtaaccctgtcgtggttcaagctgcgtgaacattaa 1140
Db 1081 TCCCTTATTAGTGTGTGTTATGATGATMCCTGCTGGGTCAAGCTGCTGACAATTAA 1140

QY 1141 gggcaagaatgaatttttcaactctgtgggttcgaagaattttagccttaaggagatga 1200
Db 1141 GGGCAAGATGAATTTTCAACTCTGTGGGTGGAAGATTTTAAAGCTTATAGGGATGA 1200

QY 1201 ttggaacctgcgaatactcgttgttcgagagatttaaggcttgatcttcccttcgtc 1260
Db 1201 TTGGAACCTGCGAATTTCTTGTTGCTGCCAGAGATTTATAGCTGTGATTTCTTGCTG 1260

QY 1261 ttgcatctcctaaccataccaacaaccttggagagcttctccttcttgaatttagca 1320
Db 1261 TTGCATTTCTGACATATCAACAACCTGGAGAGCTTTGCTCTGATTTAGATCA 1320

QY 1321 accactcgtgacatttgccttcgtatagctgtgtattctccttgggacatacaaatccttg 1380
Db 1321 ACCACTGGAACATTTGCTCTGTATGCTGTATTCCTGGGCATCACAAATACCTTTG 1380

QY 1381 ccaactatctctgaatgatgtggggcccaatcttcgaagaagtttaaccctggagaacacta 1440
Db 1381 CCACTATTTCTGGAATGATTTGGGCCCATCTTTGCCAGAAATCTTTACCTCGAGAACCTA 1440

QY 1441 ttggaagaatggcaaacgttttctgcatcgtcgtgtcatcaatgtatgttggtccatt 1500
Db 1441 TTGGAAGATGGCAAACTGTTTCTGCAATCGCTGTGATCAATGTATTTGGTCCATT 1500

QY 1501 tcttcaacatactgcgcaaaagtgtgaagtgcaaaacttgggccaatcagtgatcaccagagc 1560
Db 1501 TCTTCAACATACTGCGCAAAAGTGAAGTGCAAAATCGGCCATCAGTATCACCAAGAGAC 1560

QY 1561 acagagaatgaagaacccaataataatcctgtcctatataatgtatcttgttttcaat 1620
Db 1561 ACAGAAACTGAAGAACCAATTAATATCTGTCTATTAAATGTATCTTGTATTCAT 1620

QY 1621 gtaacctaaagtgccttataatttaagtgttaagcaatccataataagaataaat 1680
Db 1621 GTAACCTAAAGTGCCTTATATTTTAATGTAAAGTATATATACAAAGTAAATTT 1680

QY 1681 gtaactagaataattgtttagatttgaagccttgtaaatcaatgtcaatgtcactagtgc 1740

Db 1681 GTACTAGAAAAATTTGTGATTGTTGTAAGCTTGTAAATCATGAAATGCTACTAGTTGCC 1740
Oy 1741 atataagcaaaatagctatttttaattatataataaccggttgcctggaacttaattc 1800
Db 1741 ATATAGCAAAATATAGCTATTTTATATATATATTAACCGTTTGGTGGAACTTACAATTC 1800
Oy 1801 aggttcacatalctgctgcaagtcaggaagcaaccacaatlagggaggtctatattat 1860
Db 1801 AGGGTCACATATCTGGCTGCAAGTCAGGCAACCCACAATAGGGAGTCTTATTTATTTAT 1860
Oy 1861 aagaccatacccaagaatagctgtaaatagacccttctaacttggcttaataag 1920
Db 1861 AAGACCAATACCTAAAGAAATGAGCTGAAATAGACCCCTTATATACCTTTGCTTAAATTAAG 1920
Oy 1921 tggacaataattctcaagctctgtttaaacatctgttttctgaaccctcccaaaaatt 1980
Db 1921 TGGATTAATATTTCTAGGCTTTGTTAAACATCTGTTTTGTAACCTTCTCCTCAAAAAATT 1980
Oy 1981 attgtlcaacgaacatccctgacatgtagtctcaactttagcctctccacgagctgg 2040
Db 1981 ATTTGTCATCAGCAATCCCTGACATGATGATGCTCAAACTTTAGCTCTCAGGAGCTGG 2040
Oy 2041 cagccactgatatcatcagctcgtgcaacttcaactgagggagaagcagctgcc 2100
Db 2041 CAGCCACTGATATCATTCAGCCCTGGCACTTCACTGAGGGAAACATGCCAGCACCTGCC 2100
Oy 2101 acaatgccctctctgcttcaagggacagctgcccagcaacttaggagagatcccaagcag 2160
Db 2101 ACATTCCTCCCTCTCTGCGCTTCAGGGACAGTCTCCCTGGGGGCTGTTAATGTTGATGAAG 2160
Oy 2161 ggtcagcgcccaagccttgcagcgtatcttccctctggggcgtgtaactgctgtaag 2220
Db 2161 GGTGAGCCCAAGGCTTTGGAGCGATATCTTCCCTGGGGGCTGTTAATGTTGATGAAG 2220
Oy 2221 cccctgagcccaagggagcagcgcgatccacagtcagtgttccatgcaacctctccctc 2280
Db 2221 CCTTGAAGCAACAGGGAGCGCGATCCACATCATGTTGTTCCATGACACCTCTCCCTTC 2280
Oy 2281 ccttcacgacactgtagatctgctgcatgtaactctgcaaaaagaagtgtgagct 2340
Db 2281 CCTTCCAGACACTGAGATATGCTGCTGCGATGTAACCTGCAAAAAGAGTGTGCT 2340
Oy 2341 aattagcacacatalaatacatccttgaatgacatccttcaacttcaatgataa 2400
Db 2341 AATTAGCCACATATATATATCATCTTGTATGATCCTTACCTTCACATGATGATGATATAA 2400
Oy 2401 tcttcaagtctgctgttcttagagagctacacagaataataataataaagaacaa 2460
Db 2401 TCTTCAAGTCTGTTGTAGAGCTTACCAAGAAATTAATAATTAATAAAGAAACAAA 2460
Oy 2461 acattttctgctgcacacctaagtgtcgtgtagtgcagttcaaggtttaagtgaacttca 2520
Db 2461 ACATTTTCTGTCTGACACCTTAAGTGTCTGTTGCAAGTTCAAGTTAAAGGACTTCTTA 2520
Oy 2521 ctcaacataaccctgcaacccggtgtaatacatcctttagtgttcttcaaatcttat 2580
Db 2521 CTTCACATTAACCTGCAACCGGTGTGATATCATCTTATGTTGTTCTTAAATCTTAT 2580
Oy 2581 ttttccaaattttctcgtgacacatctcgaagtgttttagacatgcttggaggaccttat 2640
Db 2581 TTTTCCAAATTTTCTGGAACCATCTTCCAGTGTGTTTGAAGTGTGTTGAGGGCAATTAAT 2640
Oy 2641 gtgatttgaacctgaataatgatttcaactgtagtctcaaacacactgtaataattca 2700
Db 2641 GTGATTATGAAGCTGATTAATATGTTTCACTGTGTATGTTCAACACTGTAATATTTTA 2700
Oy 2701 actaaagctattcaatgataatgtagtgaatactctgtaaatlttttaaatctg 2760
Db 2701 ACTAAAGCTATTTAATGTAATATGATGTATACATCTGTAATTAATTTTAAATCTG 2760
Oy 2761 taaatagctttaaagtgtagtataatcttcttcaacaataaataaataatctttgg 2820
Db 2761 TAAATAGCTTTAAGTTGCTATGTGTATTTCTTTTACAAATCAAAATTAATCTTTTGG 2820

Oy 2821 aatgataaaaaaaaaaaaaaaaaa 2844
Db 2821 AATGATTAATAAAAAAAAAAAAAA 2844

RESULT 2
AX207626
LOCUS AX207626 2844 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 3 from Patent WO0156598.
ACCESSION AX207626
VERSION AX207626.1 GI:15422331
KEYWORDS
SOURCE Ovis sp.
ORGANISM Ovis sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Caprinae; Ovis.
REFERENCE 1 (bases 1 to 2844)
AUTHORS Hellerqvist, C.G.
TITLE Methods for preventing or attenuating pathoangiogenic conditions by
using the gbs-toxin (cm101) receptor as a vaccine
JOURNAL Patent: WO 0156598-A 3 09-AUG-2001;
VANDERBILT UNIVERSITY (US)
FEATURES
source location/Qualifiers
1:..2844
/organism="Ovis sp."
/db_xref="taxon:9939"
84..1571
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC60359.1"
/db_xref="GI:15422332"
translation="MKSPLVSDLAISDGEESDRPILOARAPAPVCCSARYNLAR
LSRFGEFLVLSRLVNLVSLALVDWDSNTAKDNRTSYCAEHSAPIKVLMOTKKR
WDAETLQWILDSFEFGYITITDIPGQVYASRSGKLLGFGIFATATFLTPPLADFG
VGAFLVALALELGEVGYIPMAHMSWAPLESKLSISYVGAOLGTAVSLPSG
VICYMMNTVYFVFFGIIVGILIMFLIMCLVDEPETHITIPYKEYILISLKNLSS
OKSVPMIPMLKSLPLMAIVVAHFSYNMFTYLLTLPLTYMEVLRFNIOENGLSAPV
YLQGLCMILISGDAADNLRARNSFTLVRAVRFSLIGMIGAFILYNAAGFIGDYSILA
VAFLTISITLGGFSSGFSINHLIDAPRYACILIGITRTPTFMIGPIIARSLRPE
NTIGEMQVTFICIAAINVEGAIFETLFKGEVQNWALSDHOGHRN"

BASE COUNT 745 a 623 c 594 g 882 t
ORIGIN

Query Match 100.0%; Score 2844; DB 6; Length 2844;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2844; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cccggggcgggggggttcggcggtcccgctgtagctctcttccggggagcaggtttg 60
Db 1 CCCGGGGCGGGGGGCTTCGGCGGCTCCCGCTGAGCTCTCTTTCCGCGGACGAGGTTTG 60
Oy 61 cgcgtagctcccttgaagagcatgaaagttcccggtttcggacttagcccgagcgagc 120
Db 61 CCGCGTAGCTCCCTGAAGAGCATGAAAGTCCCGGTTTCGGAGCTTAGCCCGAGGAGAG 120
Oy 121 gcgagagagggctcggagcgcacacgcgtctctgcaacgcgccccggcgggagaccgctc 180
Db 121 GCGAGAGAGGGCTCGGACCGCACCGCTCTCCACAGCGCCCCGGGGGGGAACCCGCTC 180
Oy 181 cagtagtgcctcgtctcgttcaacacttagcaatttctgctcttcttggttcttcgctc 240
Db 181 CAGTATGCTGCTCTCTCTCTTACAACTTAGCATATTTTGTGCTTTTGGTTTCTTCGTTTC 240
Oy 241 tctattcatiaacgggtgaatctgagcgcttgcactagtggaacatggatgtaataacacaa 300
Db 241 TCTATTCATTACGGGTGAATCTGAGCGTTCACACTAGTGACATGTGGATTCAAACACAA 300
Oy 301 ctggccaagaatataagaacgtctcaagagtgtagcagagcatctcgtcccaataaagctc 360
Db 301 CTGCCAAAGATTAATGAAACGTCTCAGAGTGAGAGCATTCGTCCCATTAATAAAGTTTC 360

| | | | |
|----|------|---|------|
| Db | 2521 | CTTCACATTAACCTGCAACCGGTGGTGATATCAATTTAACTGTGGTCTTTAAACTTAT | 2580 |
| Oy | 2581 | ttttccagtttttccctgagccacttccagtggttttgagcagtgttgaagccattat | 2640 |
| Db | 2581 | TTTTCCAGTTTTCCGTGACCATCTTCCAGTGGTTTGAGCATGCTTTGAGGCAATTAT | 2640 |
| Oy | 2641 | gtgattggaacttgattaaatgtttcactggtatglttcacaactcctgtaatat | 2700 |
| Db | 2641 | GGATTATTGAACCTTATTAATGTTTCACTGCTGTAATGTTCAACACTACCGTGAATATT | 2700 |
| Oy | 2701 | actaaagctatttaatgtaatatgltgtataacattctgtaaatlaattttaaactcg | 2760 |
| Db | 2701 | ACTAAAGCTATTTAATGTAATATGATATGATATCAATCTGTAAATTAATTTTAAATCG | 2760 |
| Oy | 2761 | taaatagcttaagtgtcatgtgataatttcctttaacaatcaaatlaactctttg | 2820 |
| Db | 2761 | TAAATAGCTTTAAGTGTGATATGCTATATTTCTTTACAAATCAAAATTAATCTTTTGG | 2820 |
| Oy | 2821 | aatgataaaaaaaaaaaaaaa 2844 | |
| Db | 2821 | AATGATTAATAAAAAAAAAAAAAA 2844 | |

| | | | |
|------------|-----------------------------------|-------------|-----------------|
| RESULT | 3 | | |
| AXI38494 | | | |
| LOCUS | AXI38494 | 2512 bp | DNA |
| DEFINITION | Sequence 2 from Patent EP1069184. | | linear |
| ACCESSION | AXI38494 | | PAT 30-MAY-2001 |
| VERSION | AXI38494.1 | GI:14274389 | |
| KEYWORDS | | | |
| SOURCE | human. | | |

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
I (bases 1 to 2512)

| TITLE | JOURNAL |
|---|-------------------------------------|
| Human anion transporter gene implicated in sala disease and lysosomal stalic acid transport | Patent: EP 1069184-A 2 17-JAN-2001; |

| FEATURES | Location/Qualifiers |
|------------|---|
| SOURCE | 1. .2512 /organism="Homo sapiens" /db_xref="taxon:9606" |
| BASE COUNT | 612 a 585 c 566 g 749 t |
| ORIGIN | |

| | | | | |
|----------------------------|-------|----------------|-----------|-------------|
| Query Match | 55.3% | Score 1572.4 | DB 6 | Length 2512 |
| Best Local Similarity | 82.1% | Pred. No. 0 | | |
| Matches 1917; Conservative | 0 | Mismatches 386 | Indels 33 | Gaps 8 |

| | | | |
|----|-----|--|-----|
| OY | 5 | gggggaggggaggttcggggggtccgcgcgtgagttctcttttcgcggagagaagtttgcgc | 64 |
| Db | 194 | ggcccccgggggcttgcgggctgttggccggcgctcccttcttgcacaggtgggagttacac | 253 |
| OY | 65 | gtagctccctgaaggacatcatgaatgcccggttttcgacttagcccccagacgagcga | 124 |
| Db | 254 | CT - -GCTCAGCGAGGCGTCATGAGGCTTCGCCGTTGAGACCTGGCCGGAGCATGGCGA | 311 |
| OY | 125 | ggagggcttcggaaccgcacacacgctctctgcaagcgcccgaggcggaaccgcgtccagt | 184 |
| Db | 312 | GGAGGACAGCGGACCCACGCCCTTTCMACCGGGCCCCACACGGGCGGAAGCCGCTCACT | 371 |
| OY | 185 | atcgctcgtctgtctgaaacctagcattttctgtcctttttgttttttcgtctctca | 244 |
| Db | 372 | GGGCGTCTGCTGCTTACAACTTAGCAATTTTGGCCTTTTGGTTTCTTATTTGTGA | 431 |
| OY | 245 | ttcatcaagggtgaactcgtgagcgttgacatagtgacatgltgalttcaaacacaactgc | 304 |
| Db | 432 | TCCATTAGCTGTGAATCTGAGTGTGGCCGTACTGATATGCTAGATTTCAAAATCAACTTT | 491 |

| | | | |
|----|------|---|------|
| QY | 305 | caagaataatagaaagcttccctacagatgtgtgcagaagcattctgtcccaataaaagttctca | 364 |
| Db | 492 | AGAAATATAATAACTTCCAAAGCGCTGTCCAGAGCATTTCTGCCATAAAAGTTTCAACA | 551 |
| QY | 365 | caaccaaacccggtaaaagtaaccggttggagatgacaaaaactcaagatagattctcgtac | 424 |
| Db | 552 | TAAATAAAGGGATGAAGAGTACCAATGGGATGCACAAACTCAAGATGGATTCTGGTTCC | 611 |
| QY | 425 | ttttcttcatatgctacatcatcacacaaatctccttggaagatatgtctgcagaagaatg | 484 |
| Db | 612 | CTTTTATTATGGCTATCATCATCACACACAGATTCTTGAGAGATATGTGGCAGCAAAATAG | 671 |
| QY | 485 | ggggaagcgttgtctaggagatctcgggaatcttgtcagcgtacatctcaccctgttaactcc | 544 |
| Db | 672 | GGGGAAATATGCTGCTAGAGATTGGGATCCCTTGGCACTGTCCTCCACCCGTTTCACTCC | 731 |
| QY | 545 | ccctgcctcgaatcttcggagatcggagcccttgttcacatagggacagagaaggtctgg | 604 |
| Db | 732 | CATTCTCTGAGATTATGAGGTGGAGCCACTCATTTTACTCAGAGCATGAAAGGACTGTGG | 791 |
| QY | 605 | aaagggtgcacatactcaagccatcatgcccagtgtgcttcattatgggtccccccttga | 664 |
| Db | 792 | AGAGGGTGTACATTTCAGCCATGCATGCGCATGTGGTCTTCTTGCGCTCCGCCCTTTGA | 851 |
| QY | 665 | aagaagcaagcctctcgtgatttcalatgacagagacaaacttggagacagtattctctc | 724 |
| Db | 852 | AAGAAGCAAACTCTTTACCATTTTCATATGACAGCACACAGCTTGGGACAGTAAATTCTCT | 911 |
| QY | 725 | tcctcttctcgtgagtaatttgtctactatataatgagactatgctctctatcttcttgg | 784 |
| Db | 912 | TCTCTTTCTGGAAATAAATTTGCTACTATATGATGATGGACTTATGTCTTCTCATTTTGG | 971 |
| QY | 785 | catgttgaaatcatctggttattcttlaatgatactgcgtcagttaatgataccagaac | 844 |
| Db | 972 | TACTATTGGAAATTTTGGTTCTTTTGTGGATCTGGTTAGTTAGTGTGACACACCCAAAA | 1031 |
| QY | 845 | tcacaagaacaatcacctcccatlgaataaagagatatalcttcttcacatlaaaaaatcacgt | 904 |
| Db | 1032 | ACACAAGAAATTTTCCCATTTATGAAGAAAGAAATACATCTTTTCATCATTAAGAAATCAGCT | 1091 |
| QY | 905 | ctcttcaagaagtcgaatgcgttgataacctaactgctgaaatcaatgcacatttgggtcat | 964 |
| Db | 1092 | TTCCTTCAAGAAAGTACAGTCCGTGGGTCCCACTTTTAAATTCCTCGCCACCTTTGGGCTAT | 1151 |
| QY | 965 | tgtgttcacaaatttcttcaaaccttgagacttttaacttgttgagacttaatgtccac | 1024 |
| Db | 1152 | CGTAGTGCACACTTTTCTTACACACTGCACTTTTATACTTTATTTAGCATTTATTCCTTAC | 1211 |
| QY | 1025 | ttacatagaagaagtccttaaggtlccaataltaacaagaatgggttttatactgcagctccc | 1084 |
| Db | 1212 | TTAATATGAAGGAGATCTCAAGAGTTCAATGGTTCAAGAGATGGGTTTATCTTCAATGCC | 1271 |
| QY | 1085 | ttatttagtgttgttatgtatgatactcgttcggttcaagctgtctgcaatllaagggc | 1144 |
| Db | 1272 | TTAATTAGGCTCTTGGTTATGTATATATCTCGTCTGTCAAGCTGTGCACAAATTAAAGGGC | 1331 |
| QY | 1145 | aagatggaattttcaactcgttgggttcgaagatttlaagccctlaaggaatgtatgg | 1204 |
| Db | 1332 | AAAATGGAAATTTTCAACTTATATGTGTTCGCAAGATTTTAACTTATATGGAATATTTGG | 1391 |
| QY | 1205 | acctgcgatlatcctggttgcgcgcagagatltlaagcctgtgataltctcctgtgcgtg | 1264 |
| Db | 1392 | ACCTCAGATATCCGTGAGTGCCTGGGCTCAATTTGGTATTTCTTTGGCCGCTTGC | 1451 |
| QY | 1265 | attcctaacatatacaaaacccttggagagcttctgtcctctcgtgatttagcatcaaca | 1324 |
| Db | 1452 | TTTTCTTAATATATCAACAACACTGGAGGCTTTTGTCTTCTGGATTATGATCAACCA | 1511 |
| QY | 1325 | tctggaacttgcctctgtatgacgtgatatctcctcgggacacaaataactcttggcac | 1384 |
| Db | 1512 | TCTGGAATTTGCTCTTGTATGTCTGGATCTCTCTGGGCATCACAAAATACATTTGGCAC | 1571 |
| QY | 1385 | tattccttgaaatcgttggcccatcatatgtccagaagcttlaacccttggaaacatttgg | 1444 |

QY 485 ggggaagcgtgtgtagagatctcggaatcttctgtaacagatcatcaccgtgtacacc 544
 Db 672 ggggaagcgtgtgtagagatctcggaatcttctgtaacagatcatcaccgtgtacacc 731
 QY 545 cctgcgtcagagatctcggaatcttctgtaacagatcatcaccgtgtacacc 604
 Db 732 CATGTGTGAGATTTGGAGTTGGACACTCATTTACTACAGACCTAGAAAGACTAG 791
 QY 605 agaggtgtcacatcatcaccatcgatcgatgtgtcttcaaggccccccttga 664
 Db 792 AGAGGTTTACATTTCCAGCCATGCATGCTGTCTTGGGCTCCCTCTTGA 851
 QY 665 aagaagcaagcgtgtgtagatctcatatgacagagacacactgtggagagtagtctct 724
 Db 852 AAGAAGCAAACTCTTAGATTTTCATATGAGAGACAGCTGGAGACATTTCTCT 911
 QY 725 tctcttctggaagtaattgtcactatagaaatggaactatgcttctctatcttctg 784
 Db 912 TCTCTTTTGGATTAATTTGCTACTATATGAATTTGACTATGTCTTACTTTTGG 971
 QY 785 catgtgtgaatcatctgtgttattttagatctgtcttgaatgtatgaaccagaagc 844
 Db 972 TACTATTTGGAATATTTGGTTTCTTTTGTGATCTGTGATTTAGTACACACACAAA 1031
 QY 845 tcacaagaacaatcactccctatgaaaagagataltcttcttcaacatlaaaatcagct 904
 Db 1032 ACAGACAGAGATTTTCCATTATGAAAGATATCATTTCTTCACTTAAGAAATCAGCT 1091
 QY 905 ctcttcacagaagtcagatgcgtgataaccatgctgaatcaacacagcacttggactat 964
 Db 1092 TCTCTACAGAAAGTACAGTCCGCGGTGATCCCAATTTAAATCCCTCCACTTTGGGCTAT 1151
 QY 965 tctcgttgaacatttcttcaacagagactttagtctttagtctttagtctttagtctc 1024
 Db 1152 CGTAGTGCACCTTTTCTTACACAGTGAATTTTATCTTATTTACATTTATTTGCTTAC 1211
 QY 1025 ttacatgaaaggaagtcactaagttcaataltcaagaagatgggttcttctcagctcc 1084
 Db 1212 TTTATGAGAGAGATCTCAAGGTTTCAATGTTCAGAGAAATGGGTTTATCTTCAATTTGCC 1271
 QY 1085 ttattaggtgtgtgttatgtatgtatctgtcgtggtcaagctgcgcgaacattgaaggc 1144
 Db 1272 TTTATTTAGCTCTTGGTTATGTATGATTCCTGTGTCAAGCTGACAAATTTAAGGGC 1331
 QY 1145 aagatgaaatlttcaacactgtgtgtgtcgaagatlttaagcctatagggatgttg 1204
 Db 1332 AAATGGAATTTTTCACCTTATGTGTGTGCAAAATTTTACGCTTATAGAAATGATTTGG 1391
 QY 1205 acctgcatattctcgtgtgtgcgcaagattatagagctgtgatacttctgtgtgtg 1264
 Db 1392 ACCTGCAGATTCCTGTGTGTGTGTGCTGCTTCAATTTGCTGATTTATTTTGGCCGTGG 1451
 QY 1265 attcctaacacatacaaacctctggagagcttctgtcttctggaattagacaacca 1324
 Db 1452 TTTCCCTAATATATTAACAACAACCTGGAGGCTTTTCTTTCGATTTAGCATCAACA 1511
 QY 1325 tctggaatgtgtcctctgtatgtatctcctgggacatcaaaaataaactcttgcac 1384
 Db 1512 TCTGGATATTTGCTCTTCTGTATGCTGTATCTCTGCGCATACAAAATATTTGCCAC 1571
 QY 1385 tatlcttggaatgtgtggcccatcatctgcagaagtccttaaccctggaacactatgg 1444
 Db 1572 TATTCAGAGATGTGGGCGCGTCAATGTGTAAGAGTCAACCCCGATTAACACTGTGG 1631
 QY 1445 agaatggaactgttcttctgcatgctgtctcatcatatgtatgtgtgcaatttctt 1504
 Db 1632 AGATGTGCAAAACCGTTCTATATTTCTGTCTGTATTAATGTTTTGGTCCACTTTTCTT 1691
 QY 1505 caacatctgcgcaagaagtgaatgcaaaaatgtggacatcaatgacacaaagaagaag 1564
 Db 1692 TACACTATTTCGCCAAAGGTAAGTACAAAACGTGGCTCTCAATGATCACCATGTGACACAG 1751
 QY 1565 aaactgaagaaccaaataaataatcctgtctctatataatgtatcttgttatacatgtaa 1624

Db 1752 ACACCTAAGGAACCAATTAATATCTCCCTCTATTAATGATTTTATTTATTCATGTA 1811
 QY 1625 cctaaagtgcttctgtatattttaaagtgaagcaatcatatacaagaataaattgtac 1684
 Db 1812 CTTCAAGGCTCTT-----CTGTATTTGTGAACCAATTTCTATGT-----CTTTT 1857
 QY 1685 tagaaatctgtgtatgttgaagctgtgtatcatatgaatgtaactagtgcatat 1744
 Db 1858 AATTTACTTTGTATTTAATTTTAAAGCCTATATATCATGAATATCATCTAGTTGCCAGAA 1917
 QY 1745 aagcaaatctgtatattttaaattatlaaaccgttctgtggaacttacaatlaaggg 1804
 Db 1918 TAATAAATGAAGACTGTGTTAATTAATGAATATATGATTAAGCTAGACTCTACTTAAAGT 1977
 QY 1805 tcaatatctgtctgcaagtcaggaacccaacataagggagttctattta-cttaag 1863
 Db 1978 TCACATACCTGCTGTGTACGTGGCAACATGAAGTGGACAGTCTGTGATTTTAAAG 2037
 QY 1864 accatacctaagaagatgagatgaaatagaacctctatcaacttcttgaatlaaggtg 1923
 Db 2038 GCATATCTAAGGAGATGAGCTGAACACAGCTCTGATACCTTTGCTTATATTAACATAG 2097
 QY 1924 ataataatctcagctctgtgtttaaacaatctgttcttgaacacttccctcaaaaatlat 1983
 Db 2098 ATGATATTTCTCAGGTACGTATTAACACCGTGTGTTCACTTCTCATTAATA----- 2152
 QY 1984 tgtcatcagaatcccttgatcatgtatgtctcaaaactttagcctctcgaaggagctgca 2043
 Db 2153 -ATTGTACGCTCTCTGTACACTTGAACCTCAAACTTGAACATCTGTGTGACCTGCAT 2211
 QY 2044 ccaatgtatcatcagcctcgtggaacacttcaatgaggaagaatgccagagctgtgcaca 2103
 Db 2212 CCACGTATATTTTCCCTGCTGCACTGAGTGAAGGAGTGTGCCAGGACGTGCCAAG 2271
 QY 2104 tgtccctctctgtgtcttcaaggagacagtgccagcaact-----aggcaacatcaaga 2156
 Db 2272 CACTCCCTCCCTGCTTCAAGGCTCAAGAGTCCAGGCTTATCAGAGGACGATCCAAGC 2331
 QY 2157 cgaaggtcagcgcgaagagctttagaaggtatcttccccc-aggagctgtatgtgtg 2214
 Db 2332 CCAGACCCAGTGTGACCTCTGCTGTGTGCTTCTCTGAGAGGCTATCAATGTGTAG 2291
 QY 2215 atgaagccctgagcgaacagggagacgcgcgcatccacagtcattgttccatgcacctc- 2273
 Db 2392 ATTAAGCCCGTAGTAGCAAGACAGACAGATCCACTGCTATGTGTATACATCTCTCA 2451
 QY 2274 tcccttcccttccagcaacactgtgaglatgtcctgtgcatgtaaactgtcaaaagaa 2329
 Db 2452 AACTTTCCCTCCACACAGAGGAATATTTGCTGGCATGCAACCTGCAGAAAGAA 2507
 RESULT 5
 AX207624
 LOCUS AX207624 2930 bp DNA linear PAT 31-AUG-2001
 DEFINITION Sequence 1 from Patent WO0156598.
 ACCESSION AX207624
 VERSION AX207624.1 GI:15422329
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 2930)
 AUTHORS Helleryqvist,C.G.
 TITLE Methods for preventing or attenuating pathoangiogenic conditions by
 using the gbs-toxin (cm101) receptor as a vaccine
 JOURNAL Patent: WO 0156598-A 1 09-AUG-2001;
 VANDERBILT UNIVERSITY (US)
 FEATURES
 source Location/Qualifiers
 1..2930
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

| | | |
|--|------|--|
| CDS | | 263..1873 |
| | | /note="unnamed protein product" |
| | | /codon_start=1 |
| | | /protein_id="CAC60358.1" |
| | | /db_xref="GI:15422330" |
| | | /translation="MAAGAMTPPPRPVQPARPGFGISGRSLICQVASTPAHWKMS |
| | | VPVDLARDDEESTDRTPLPGAPRAEAAPVCCSARNLILAFFGFIVYALRVNIS |
| | | VALVDVDSNTLEDNRTSKAPEHSAPIVHNQNKGYQMDAETQWILGSFYG |
| | | ITIQIPGVYASKIGKMLGFIIGLAVTLFPIADLGVPLVLRLEGEGEV |
| | | TFPAMHAMSSMAPLERSKLSISYAOLGVISLPSGIICYYNMWVYFPGT |
| | | IGTFWLLIMIVSDPKRKRISHKEYEILISLRQDSQKSVPRVPLKSLPLMA |
| | | IVVAHSTYNTFTLLFPTLPTMKEILRFNVQENGFLSLPYLSWMLSLQDAAN |
| | | LRAKWNSTLCVRIFSLGIMIGPVLVAGVIGCDYSLAVALFISTLGGFCSSG |
| | | FSINHDIDAPSYAGILGITNFEATIPGMVPIAKSLTPDNTVGEQVFIYAAAIN |
| | | VFGAIFETFLAKEVQWMLNDHGHHRH" |
| BASE COUNT | | 729 a 698 c 681 g 822 t |
| ORIGIN | | |
| Query Match | | 55.2%; Score 1570.8; DB 6; Length 2930; |
| Best Local Similarity | | 82.0%; Pred. No. 0; |
| Matches 1916; Conservative 0; Mismatches 387; Indels 33; Gaps 8; | | |
| QY | 5 | gggggggggggctcgccggtcccgctggagctctcttccgsgaagagttgcgc 64 |
| DB | 309 | GGCCGGGGGGCTTGGCGCTGCGGGCGGGCGCTCCCTCTCTGCCAGGTGCGAGTAC 368 |
| QY | 65 | gtagctccctaaagcatatgaagtcoccggttcgagtaagcccgagcgagcgga 124 |
| DB | 369 | CT--GCTACAGTACGCGTCAATGAGTGTCCGGTTCGAGACTGGCCCGAGACATGGCGA 426 |
| QY | 125 | ggagggctcggaacgcgaacacgcctcctcgacgcgcgcgcggcgaggaacccgctcagt 184 |
| DB | 427 | GGAAGACACGACGCGACGCGCTTCTTACCGGGCGGCCCGACGAGCGCAAGCGCTCAGT 486 |
| QY | 185 | atgttgtctgtctggttacaacctaagatcttctcctcttcttcttctcgttcccta 244 |
| DB | 487 | GTCGTGCTCTGCTCGTTACCACTTGAATTTGGCCCTTTTGGTTCTTCATTGAGTA 546 |
| QY | 245 | tctatcaggggtatctcgagcgttgcactagtgacatgtgagatccaacacatcgc 304 |
| DB | 547 | TGCAATTACGTCGATCTGAGTGTGGCTTACGTGATGATGATGATTCCTCAATACAACTTT 606 |
| QY | 305 | caaaagataagaacgctcctactagagtgacagagcatctgtcccaataaagttcttca 364 |
| DB | 607 | AGAAGATAATGAGACTTCCAAGGCGGTGCCAGACATTTGCTCCATTAAGAATGATCA 666 |
| QY | 365 | caaccacaaggggtataaagtaacgcgtgtgagatgcagaactcaagatgagatctcgatc 424 |
| DB | 667 | TAAATCAACGCGTAAGAGTACCAATGGGATGCAAGAACTCAAGGATGGAATTCGCGTTC 726 |
| QY | 425 | ttttctatgactacatcacaacaattcctcgagagatagtgtccagcagaagttg 484 |
| DB | 727 | CTTTTATATGCTACATCATCACAGATTTCCGTGAGGATATTTTCCGACCAAAATAGG 786 |
| QY | 485 | gggggaagcgtgtcgagatccggatcttgcacagctatcttccacctgtctactcc 544 |
| DB | 787 | GGGGAATAATCTGCTAGGATTTGGGATCTTGGCATGCTGTCTCCACCTGTGTACTCC 846 |
| QY | 545 | ccctgcgcagatcttggaatcgagaccctgtgtgcactaaggcactagaagggctagg 604 |
| DB | 847 | CATTGCTGCAAGATTTAGAGAGTGGACCACTCATTTGTACTAGACACTTAAGAGACTAGG 906 |
| QY | 605 | agaaggtcacatcatccagcatgcatgtgattgtcttcaatggggtccccccttga 664 |
| DB | 907 | AGAGGGGTTCATATTTCCAGCATGCAATGCCATGTGCTTTCTTGGGCTCCCTCTTTGA 966 |
| QY | 665 | aagaagaacgtcttgatattcaataatgcaaggacaacttgggaacagtaatttctct 724 |
| DB | 967 | AAGAAGCAAACTCTTATGATTCGTATGCAAGAGACAGACTTGGGACAGATAATTTCTCT 1026 |
| QY | 725 | tcctctcttcgagtaattgtactataatgaattgacttaatttcttatttcttgg 784 |
| DB | 1027 | TCCCTTTCTGGAAATAATTTGCTACTATATGCAATTTGACCTTATGCTTTTGGG 1086 |
| QY | 785 | catgttggaatcatctggttatttattatgacatcgtctagtagatcaccaagaac 844 |
| DB | 1087 | TACTATTGGAATATTTTGGTTTCTTTTGTGATCTGTTAGTATGACACACCACAAA 1146 |
| QY | 845 | tcacaagaacatcactccctcctatgaagaagatattcttcttcatcatlaaanaacagt 904 |
| DB | 1147 | ACACAAGAGATTTTCCCATATGAAAAGGAAATACATTTCTTCATCATTAAGAAATACGCT 1206 |
| QY | 905 | ctctccagaagtcagtgccgtgagatacctabagctgaatcaactgcacttgggctat 964 |
| DB | 1207 | TTCTTTCACAGAACGACATGGCGGTGGACCATTTTAAATTCCTGCCACTTTGGGGTAT 1266 |
| QY | 965 | tgctgtgcacatttcttctcaactgcagcttcttcaacttcttcttcttcttcttct 1024 |
| DB | 1267 | CGTAGTTGCACACTTTTCTTACACTGACCTTTTATATCTTTATTTGACATTTAGCCTTAC 1326 |
| QY | 1025 | ttacatgaaggaagtccttaagttcaatlatcaagaagaatgggttttctatctgcagtc 1084 |
| DB | 1327 | TTATATGAGAGAGATCCTTAAGGTTCATATGTTCAAGAGATGGGTTTTATCTTCATTTGCC 1386 |
| QY | 1085 | ttatttagtgtgtgtatgtagatccgtgtgggttcaagctgcgtcaaatlaaaggc 1144 |
| DB | 1387 | TTATTTAGGCTCTTGGTTATGATGATCCCTGTGCTCAAGCTGCTACAAATTTAAAGGC 1446 |
| QY | 1145 | aagaatgaatttctcaactcgtgtgttcgaagaagtttcttaagccttaagaagatagtc 1204 |
| DB | 1447 | AAATGGAATTTTTCACACTTTATGTGTTCCAGAAATTTTATAGCTTATAGATGATGG 1506 |
| QY | 1205 | aactgcgaattccctgtgtccgcgaagatltatagctgtgatbatccttggcgtgtgc 1264 |
| DB | 1507 | ACGTGAGTATTCCTGCTACTGCTGCTGCTCATTTGCTGATATTTCTTTGGCCGCTTGC 1566 |
| QY | 1265 | attcctacacatcaacaacaccccggaagccttttgccttctcggatttagccaacca 1324 |
| DB | 1567 | TTTCTTACATATATCAACAACACTGGAGGCTTTTGTCTTCTGTGATTTAGCATCAACA 1626 |
| QY | 1325 | tctgacatctcctcctcgtatcgtgtatctcctcgtgacatacaaatccttctgcac 1384 |
| DB | 1627 | TCTGGATATTGCTCTCTGATGCTGTGATGCTGTATCTCTGCGATCACAATATCATTGGCAC 1686 |
| QY | 1385 | tattcctgaatgatttgggcccacatctgcagaagatcttacccttgagaacactatgtg 1444 |
| DB | 1687 | TATTCAGAGATGTTGGCCCGCTCATTTGCTAAAGCTGACCCCTGATTAACACTGTTGG 1746 |
| QY | 1445 | agaatgcaaacgcttcttgcacatcgtctgtcctacatagatatttgggtccatttctt 1504 |
| DB | 1747 | AGAATGGCAAAACCGTGTATATATGCTGTGCTATTAAATGTTTGGTCCATTTTCTT 1806 |
| QY | 1505 | cacactatcgcacaaggtgaagtgcacaactggccatcagtgatcaccaagagacag 1564 |
| DB | 1807 | TACACTTTTGGCCAAAGGTAAAGTACAAACTGGGCTCTCAATGATCATCCATGGACACAG 1666 |
| QY | 1565 | aaactgaaggaaccaataaataatcctgtctctatlaaagtatcttgyttlatcatgtaa 1624 |
| DB | 1867 | ACACTGAAGGAACCAATAAATAATCTGCTCTATTAATGATTAATTTATATCATGATA 1926 |
| QY | 1625 | cttaaaagtgccttgaatlatttaatgttgaacaactatatacaagaataaattgtac 1684 |
| DB | 1927 | CTTCAAAAGTGCCTT-----CTGATTTGTGTAAGCAATTTATGATG-----CTTTT 1972 |
| QY | 1685 | tagaanaatgtgttagatctgaagcctgttaactcatgaatgcactagctgcacat 1744 |
| DB | 1973 | AATTGTACTGTATATGATTTTAAAGCCCTATATATCATATGAAATATCATAGTTGCCAGAA 2032 |
| QY | 1745 | aagcaaatagctatatttaattatataaccggtttgcgtgaacttaacatlaaggg 1804 |
| DB | 2033 | TAAATAAATGACGTGTTTAAATTATGAATAATATGTAAGACTGTGAGACTTCTACTTTAGT 2092 |
| QY | 1805 | tcacatatctgctgcgaagtcaggaaccccaataagggaggttctattta-cttataag 1863 |
| DB | 2093 | TCAATATACCTGCTGCTAGTCGGGCAACATGAAGTAGACAGTTCTGTGATTTTATAG 2152 |

| | | | | | | | | |
|----------------------------|---|---|-----|---|-----|---|-----|---|
| BASE COUNT | 729 | a | 698 | c | 681 | g | 822 | t |
| ORIGIN | | | | | | | | |
| Query Match | 55.2%: Score 1570.8; DB 9; Length 2930; | | | | | | | |
| Best Local Similarity | 82.0%: Pred. No. 0; | | | | | | | |
| Matches 1916; Conservative | 0; Mismatches 387; Indels 33; Gaps 8; | | | | | | | |
| QY | 5 | ggggcgggggcgcttcgcgcgtccgcgttgaaagctctcttccgcggaagcaggttcgcgc | 64 | | | | | |
| Db | 309 | GGCCCGGGGGCGCTGGGGCGTGTGGGGCGGGCGCTCCCTTCTGTGCCAGGTGGCGAGTAAAC | 368 | | | | | |
| QY | 65 | gtagctccctgaaggaatcatgaatgctcccggttcggaactatagcccgagacgcgcga | 124 | | | | | |
| Db | 369 | CT--GCTTACGCTAAGCGCATATGAGGTCTCCGAGTTTGGAAGACTTGCGCCGGAACGATGGGGA | 426 | | | | | |
| QY | 125 | ggaagagctcggaaacggcaaccgcgtctcggcagcgcggcccgcgggcggaaccgcgtccagt | 184 | | | | | |
| Db | 427 | GGAGAGCACGGGACGCCACGCCCTCTTCTTACCGGGGCCGCCACAGGGCCGAAGCCGCTCCAGT | 486 | | | | | |
| QY | 185 | atgcgtcctcgtcgttacaactacaactgacatttgcctcttlltggtttcttcgtctctca | 244 | | | | | |
| Db | 487 | GTGCGCTGCTGCTGTTTCAACTATGCAATTTTGGCCCTTTTGGTTTCTTCAATTGTGTGA | 546 | | | | | |
| QY | 245 | ttcatcagggatgaatcgtgacgcttcgacatgtggaatcgtgtgaaatcgaacacaactgc | 304 | | | | | |

| | | | |
|----|------|---|------|
| QY | 305 | caagatataagaagcttctcagagctgycgaagagattctgtcccatataaagttcttca | 364 |
| | | | |
| Db | 607 | AGAAATATATAGAACTTCCACAGGCGTGTCCACAGAGATTCTGCTCCATPAAAGTTTCATCA | 666 |
| | | | |
| QY | 365 | caaccaaacgggtaaaaagtaaccggttgagtcgaagactcaaggtatgatctcggatc | 424 |
| | | | |
| Db | 667 | TATATCAACGGGTAAAGAGTACCAATGGGATTCAGAAACTCAAGATGATTTCTGGTTTC | 726 |
| | | | |
| QY | 425 | tctttctatgctacatcatcaacaactctccttgagagatatgttcgcagagaagtgg | 484 |
| | | | |
| Db | 727 | CTTTTATATGGCTCATCATCACACAGATTTCCTGGAGATATATGTTGCCAGCAAAATAGG | 786 |
| | | | |
| QY | 485 | ggggaagcttgttctagagattcggagatctttgtctacagatcttcaccctgttaacctc | 544 |
| | | | |
| Db | 787 | GGGGAAATATGCTGTGATGATTGGATCTTGCGACTCTGTCCACCTGTTCCATCC | 846 |
| | | | |
| QY | 545 | ctctgcctcagaatttcggaatcggagccctgtgtacatcaggagactagaagggctagg | 604 |
| | | | |
| Db | 847 | CATTGCTCAGATTTTAGAGATTGGACCACTCATTTTACTCAGAGACATAGAAAGGACTGAG | 906 |
| | | | |
| QY | 605 | agaagggtgcacatataccaagcatatgacatgctatgtgtctcatatgagcttccccccttga | 664 |
| | | | |
| Db | 907 | AGAGGGTTTACATTTTCCAGCCATGCATGCGCATGGTCTTCTTGGGCTCCCCCTCTTGA | 966 |
| | | | |
| QY | 665 | aagaagaagcttctgagatctcatatcatatgcaaggagacaaacttggagacagtagttctct | 724 |
| | | | |
| Db | 967 | AAGAAGCAAACTTCTTACCATTTTCTATGACAGGACACAGCTTGGGACAGTAATTTCTCT | 1026 |
| | | | |
| QY | 725 | tccctcttcggaagaaatttgcataataataaattggagctatgctctctatctctctgg | 784 |
| | | | |
| Db | 1027 | TCTCTTCTTCTGGAAATAATTTGCTACTATAATGAAATTTGGACTATATGCTTCTACTTTTTTGG | 1086 |
| | | | |
| QY | 785 | catgtctgaaatcatcctgttatttattttagatctgctatgttagtga taaccaagaac | 844 |
| | | | |
| Db | 1087 | TACTATTGGAAATATTTGGTTCTTGTGTGGATGCTGTTAGTAGACACACACAAAA | 1146 |
| | | | |
| QY | 845 | tcacagaagcaatcacctcctatgaaagagtagtatatttctcatcatataaaataacagct | 904 |
| | | | |

Db 1147 ACACAGAGAAATTCACATTAATGAAAGAAATACATTCCTTACATTAATGAAATACAGCT 1206
 QY cctccacagaagtcagtcgctgagataccatgctgaatccactgcccacttgagctat 964
 Db 1207 TTCTTCACAGAGAAATGCGCGGTGATACCATTTTAAATCCCTGCCACTTTGGGCTAT 1266
 QY 965 tgcctgtagacatttctctacacacggaacttttatacttctgtagacttgcctac 1024
 Db 1267 CGTGTGGACACTTTCTTCAACTGACCTTTTATATACATTAATGACATTAATGCGCTAC 1326
 QY 1025 ttacataagaagaagctgaagctgaatcatatcaagaagaatgagctttatctgagcc 1084
 Db 1327 TTATATGAGAGAGATCCATAGGTCAATGTTCAAGAGAGATGGTTTATCTTATATGCGC 1386
 QY 1085 ttattagtgctgtagatgatgatccgctgcggtcaagctgcagcaacttaagagc 1144
 Db 1387 TTATTTAGGCTCTGGTTATGATGATCCTGTCGTCACACTCTGACATTTTAAGGC 1446
 QY 1145 aagatggaatttctcaactctgaggtcgagaagcttttagacctabagagatgattg 1204
 Db 1447 AAAATGGAATTTTCAACTTATATGTTGCGAGAAATTTTGGCCTTATAGGAATGATTGG 1506
 QY 1205 acccggaatctcctgctgctgcgaggaattatagctgtagatcttccctgctgttc 1264
 Db 1307 ACCTGACGATATCTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1566
 QY 1265 attcctaaccatacaacaaccccgagagccttctcctcctcctgattagcatcaaca 1324
 Db 1567 TTCTCTACATTAATCAACACACTGGAGGCTTTGCTCTTCTGATTTAGCATCAACCA 1626
 QY 1325 tctggaactgctcctcctgtagctgtagtctcctcctcctcctcctcctcctcctc 1384
 Db 1627 TCTGGATATCTCTCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1686
 QY 1385 tattccgggaatgctgagcccatctgccaagaagcttcaacccctgagaacactattg 1444
 Db 1687 TATTCAGGAATGTTGGGCCGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1746
 QY 1445 agaattgcaaaccttctcctgtagcctgctcctcctcctcctcctcctcctcctc 1504
 Db 1747 AGATGGAACCAACCGTGTCTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1806
 QY 1505 cacacattcgcgaagaagtggaagtgcaaaactgagccatcagtgatcaacaagaagacag 1564
 Db 1807 TACACTATTCGCCCAAGAGTGAAGTCAAAACTGGGCTCTCAATGATCAACATGAGACAG 1866
 QY 1565 aaactggaagaacaataaataatcctgctcctcctcctcctcctcctcctcctcctc 1624
 Db 1867 ACCTGGAAGGACCAATTAATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1926
 QY 1625 cctaagaagtccttctgtagatcttctgaagcaatcctcctcctcctcctcctcctc 1684
 Db 1927 CCTCAAGTGGCTT-----CTGATTTGTTGTAAGCATTTCTATGT-----CTTTT 1972
 QY 1685 tagaanaattgctgtagatttgtaagctgtaacatgaatgctagctagtgccat 1744
 Db 1973 AATGTACTCTGATTAATTAATTTTAAGGCTATATCATGAAATATCACTAGTTCAGAA 2032
 QY 1745 aagaanaattgctatttataatataataaccgcttctgtagaacttataactcagg 1804
 Db 2033 TTAATTAATGAAGATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2092
 QY 1805 tcaacatactcgtcgaagtcgaagcaaccacaaatagaggagcttcttcttcttcttct 1863
 Db 2093 TCACATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2152
 QY 1864 accatacctaagaagtcgaagtcgaagcaaccctcctcctcctcctcctcctcctcctc 1923
 Db 2153 GCCATACTAAAGGAATGAGCTGAACAGACACTCTGATACCTTTGCTTAAATTAAGTAG 2212
 QY 1924 ataataattctcaggtcttgtaaacatctgcttcttcttctcactcctcctcaaaatatt 1983
 Db 2213 ATGATATTTCTCAGGTACTGATAAAGACCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2267

QY 1984 tgcatacgaatccctcctgacatgtagtctcaaacctttagcctcctcctcagagctgag 2043
 Db 2268 -ATTGTCAGCTCTCTGACACTTAGACCTCAACTTATGATCTGCTGAGCTGGCAT 2326
 QY 2044 ccactgataatcagctcctggaactcctcctcctcctcctcctcctcctcctcctcctc 2103
 Db 2327 CCACCTGATATATTTTCGCTGCACTGACCTGAGAGGAGTGTGCTCCAGCAGCTGCCAAG 2386
 QY 2104 tgcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 2446
 Db 2387 CACTCCCTCCCTGCTTCAAGGCTGAGAGTCCAGCTTATATCAGAGCAGCATCCAGC 2446
 QY 2157 ccaggtcagcgcgaagccttgcaggtatcttccctc--ggggtctgtaagtgtg 2214
 Db 2447 CCAGAGCAGAGTGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2506
 QY 2215 atgaagcctcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagc 2273
 Db 2507 ATAAAGCCCTGAGTAGAGCAGAGCAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTG 2566
 QY 2274 tccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 2329
 Db 2567 AACTTCCCTTCCAGCAGAGAGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2622

RESULT 7
 BC020961
 LOCUS 3329 bp mRNA linear PRI 22-JAN-2002
 DEFINITION Homo sapiens, soluble carrier family 17 (unilor/sugar transporter), member 5, clone MGC:8885 IMAGE:3847279, mRNA, complete cds.
 ACCESSION BC020961
 VERSION BC020961.1 GI:18088702
 KEYWORDS MGC.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 3329)
 AUTHORS Strausberg, R.
 JOURNAL Direct Submission
 Submitted (03-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgaps-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www.shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 20 Row: 1 Column: 2
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6912665.
 FEATURES
 source
 1..3329
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:8885 IMAGE:3847279"
 /tissue_type="Colon, adenocarcinoma"
 /clone_lib="NIH_MGC_65"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

| | |
|----------------------------|--|
| CDS | 125. .1612 /codon_start=1 /product="solute carrier family 17 (anion/sugar transporter), member 5" /protein_id="AAH20961.1" /db_xref="gi:18088703" /translation="MRSFPRDLARNDESESDRPLPGAPRAEAAPVCSARYLAI LAFGEFLVAVRLNLSVALVDMSNTLEDNRSKAPRHSAPLKYVHNHDTGKKYQ WDAPGQILGSEFEGYLIIDIGSVASKSGKMLGFLGTRAVLTFPIADLG VGPLIVALEBGLSEGVTFPPMHAMSSWAPLESKILSTIGAQLGTVISLPLSG LICVMNTVYFEFFGTIGIFWFLMWLVDTPQKRIISHYEKEYLISLRQLSS QKSVWPILKSLPLMAIVAHFSTNMTFYLILPLPYMEILRFENQENGFSLP YLGSWLMILSGAADNLRAKNFSTLCVRILFSLIGIGIAVFLVAGFTGCDYSLA VAPLISITLLGFCSSGSINHLIDAPVAGILIGITNPTFTIGMWPVAKSLITPD MTVGEWQVFIFIAAINFGAIFPLTPKQEVQWALMDHHGHRH" |
| BASE COUNT | 863 a 718 c 727 g 1021 t |
| ORIGIN | |
| Query Match | 55.2%; Score 1570.8; DB 9; Length 3329; |
| Best Local Similarity | 82.0%; Pred. No. 0; Mismatches 387; Indels 33; Gaps 8; |
| Matches 1916; Conservative | 0; Mismatches 387; Indels 33; Gaps 8; |
| OY | 5 ggagggggggggttcgagcgtcccgctgagctctctcttccgcgagcaggttgcgc 64 |
| DB | 48 GGGCCGGGGGCTTCGGGCGTCTGGGGCGCGCTCCCTTCTGCGCAGTGGGAGTAC 107 |
| OY | 65 gtagctccctgaagacatagaagtcccggtttcgaactaagcccgagcgaagcga 124 |
| DB | 108 CT--GCTACAGTAGGCGTCAATAGGTCTCCGGTTCGAGACTGCGCGGAGCGGCA 165 |
| OY | 125 ggaaggtcgcgacgcacacgcgctcctgacgcgcgcgcgcgcgcgcgcgcgcact 184 |
| DB | 166 GAGAGACGACGACCGCAGCCCTCTTCTACCGGGCGCCCGCCGAGGCGGCTCCACT 225 |
| OY | 185 atgctgctctgctgttaacaacctagcaattgttccttlttgglttcttcctcta 244 |
| DB | 226 GTCGTCGTCGTCGTTACACTTAGCAATTTGGCCTTTTGGTTCTTCATTGTGTA 285 |
| OY | 245 ttcatlaagcgtggaatcgaagcgttgcaactagtagaactggtggaatcaacaacgcgc 304 |
| DB | 286 TGCATTACCTGTGAATCTGAGTGTGCGGTAGTGTGATAGTATTCATAAACAACCTTT 345 |
| OY | 305 caaagataataagaactccacagtgtagagcaacttctgctccataaagaactctca 364 |
| DB | 346 ACAAGATATATACCTTCCAAGGCAATGTCCAGACATTTGCTCCATTAAGTTCAITCA 405 |
| OY | 365 caaccaaacggttaaaaagtlacccggtlbggaatgcagaacactcaaggatctcgcgat 424 |
| DB | 406 TAATCAACGCGGTAGGAAGTATCAATGGGATGAGAACTCAAGATGTCTCGGTTTC 465 |
| OY | 425 ttttttctatgctacatcaacaacaaatccctggagagatagtgtccagcagaagaatg 484 |
| DB | 466 CTTTTTTTATAGGCTACATCATCAGACAGATTCCTGAGAGATATGTGTGCAGCAAAATATAG 525 |
| OY | 485 ggggaagcgtgttcgaagatcgcgagctcttgtaacgactatctcaaccgcttcaacc 544 |
| DB | 526 GGGGAAAAAGCTGCTAGAGATTGGGATCTTGACACTGCTGTCTCCACCCCTTTCACTTC 585 |
| OY | 545 cctgcgcgcgaatttcgagatcgcgacccctgttgcaactcaagcgaactagaaggatcga 604 |
| DB | 586 CATTGCTGCAGATTAGGAGTTGAGCACTCATTTGACTCAGAGCACTAAGAGAGACTAGG 645 |
| OY | 605 agaaggtgtcaacatccagccaatgcagatgtgtgtcttcaatgaggtcccccctctga 664 |
| DB | 646 AAGAGGTGTATTAATTCAGCCATCATGCCATGTGTGTTCTTGGGCTCCCTCTCTTCA 705 |
| OY | 665 aagaagcaagctctgagatctcatatgacagagacaaacttggagacagtaattctct 724 |
| DB | 706 AAGAAACAAACTTCTTAGCATTTCAATGACAGACACAGCTTGGAGACAGTAATTTCTCT 765 |
| OY | 725 tccctcttcctgagaaattgtctactatagaattggaacttgcctctatattcttgg 784 |
| DB | 766 TCCCTCTTCTGAAATTAATTTGCTACTATATGAAATTCGACTATGCTCTACTTTTTGG 825 |
| OY | 785 ccttgttgaatcactcctgttattattatgatactcgttagtgaatgaacccgaagac 844 |
| DB | 826 TACTATTTGGAATATTTTGGTTCTTTTGGATCGGTAGTATGAGACACACACAAAA 885 |
| OY | 845 tcaagaacaaactccctccctgaagaagagatatctcttcaatcaatlaaanaatcagct 904 |
| DB | 886 ACACAGAGAAATTTCCCATTAATGAAAGAGAAATACATTTCTTATATCTTAAGAAATCAGCT 945 |
| OY | 905 cctctcacaagaatcagtcgcgttggaatacctatgctgaaatacactgcccacttgggctat 964 |
| DB | 946 TTTCTACAGAGATCAGTCCGCGGTACCCATTTAAATCCCGCCACTTTGGGCTAT 1005 |
| OY | 965 tctgcttgcacatcttcttcaaaccttgacttltatcttgttgacttatgtcctac 1024 |
| DB | 1006 CGTAGTTCGACACTTTTCTTACACGCACTTTTATATCTTATGACATTTATGCTTAC 1065 |
| OY | 1025 ttacatgaagaagctccctgaagttcaatactcaagaagaatgggttlttactgcagctcc 1084 |
| DB | 1066 TTTATGAAAGAGATCCTTAGGTTCAATGTTCAAGAGAAATGGGTTTTTATCTTCAATGGCC 1125 |
| OY | 1085 ttatttgaatgttggttataatgatactcgtcgcgttcaagctgcgtcaaatlttaagggc 1144 |
| DB | 1126 TTTATTAGGCTCTGCTGTTATGATGATCCTGCTGCTGCAAGCTGTCGACAAATTTAAGGCG 1185 |
| OY | 1145 aagatggaatttctcaactctgtgggttcgaagaagttttaaagcctataaggatgattg 1204 |
| DB | 1186 AAAATGGAATTTTTCACATTTATGTGTGCGAGAAATTTTATACCTTATGAGATGATTTGG 1245 |
| OY | 1205 accctgatatctcgtgttgcgcgaagatlttaagcgttgatacttctcgttcgttcgt 1264 |
| DB | 1246 ACCTGCAATTCCTGATGCTGATGCTGCTGCTTCAATTTGGCTGATTTATCTTTGGCGTTTC 1305 |
| OY | 1265 attcctaacatalcaacaacccctgggaagcttctgctcctcttggaattagaatcaacca 1324 |
| DB | 1306 TTTCTTAATATATATCAACAACTGAGAGCTTTTGTCTCTCGATTTAGCATCAACCA 1365 |
| OY | 1325 tctggaactgtcctccttgatagctggtatcttctcgtggacatcaaaactcttggcac 1384 |
| DB | 1366 TCTGATATGTCCTGCTGATGCTGATGCTCCTGGGCATCAACAAATATTTGCCAC 1425 |
| OY | 1385 tatctcgaatgatatgggcccacatctgcagaagcttcaaccctgagaaacactatg 1444 |
| DB | 1426 TATTCAGGAATATGTTGGCCCGGTCAATTCGTAAGATCTGACCCCTGATTAACACTGTGG 1485 |
| OY | 1445 agaaltggcaaacgttctctgcacgcgtgcgtacatcaatglatattgtygcacttctt 1504 |
| DB | 1486 ACAATGCGCAAAACCGTGTCTATATGCTGCTCTATTAATGATTTTGGTGGCATTTTCTT 1545 |
| OY | 1505 caactattcgcgaagaagtgaagtgcgaacactgggcatcagtgatcaacgaagacag 1564 |
| DB | 1546 TACACTATTTGCGCAAAAGGAGTACAAAGCTGAGGCTCTCAATGATCACCAATGAGACAG 1605 |
| OY | 1565 aaactgaaggaacaaataataatcctgctctatlaaatgtaacttcttggltcaatgtaa 1624 |
| DB | 1606 ACACGGAAGGACCAATTAATATATCTGCTCTATTAATGATTTTATTTATTTATTCATGTA 1665 |
| OY | 1625 cctaaagtgcccttgataattlaatgltgtaagcaactcatataacaagaataatgttac 1684 |
| DB | 1666 COTCAAAAGGCTT-----CTGTATTTGTAAGCAATTCATGT-----CTTTTCTT 1711 |
| OY | 1685 tagaaanaattggttagatttgaagcctgttaatcaagaatgtcactagttgcacat 1744 |
| DB | 1712 AATTTACTTGTATTAATTTTAAAGGCTATTAATCAATGAAATATACATGATTTGCGACAA 1771 |
| OY | 1745 aagcaaatatgacttcttlaattlaattlaacccgttgcgttgcgaactcaataacaggg 1804 |
| DB | 1772 TATATTAATGAACCTGTGTTTAATTTATGATTAATATGTAAGCTAAGACTTCTTATAGCT 1831 |
| OY | 1805 tcaacatctgctgcgaagtcagcgaaccacaataagggaggtctcatatta-ctataag 1863 |
| DB | 1832 TCACATACCTGCTGCTAGTGGGCAACATGAATGAGCACTGCTGTTGATTTTGG 1891 |

| | | | |
|----|------|---|------|
| OY | 1864 | accacactaaagagatbatgctgaatagaccctctctaaccttggcttaataaggyg | 1923 |
| | | | |
| Db | 1852 | gccatcactpaaaggaaatgctgaaacagacctcctgattaccttggcttaataaactag | 1951 |
| | | | |
| OY | 1934 | ataaatctctaaagtccttggttaaacactcgtttctgtacacctctcccaaaaattalt | 1983 |
| | | | |
| Db | 1952 | atgataattttcaggtactgatatpaaacacctgtgtgtgttcaactttccatpaaaa | 2006 |
| | | | |
| OY | 1984 | tgtcatcagcaatccctctacatgtatgtgtctcaactttaagctctccacagagcttgcag | 2043 |
| | | | |
| Db | 2007 | attttgcagcttctctctgacaccttggacctcaaatpattagacattctgtggagctccat | 2065 |
| | | | |
| OY | 2044 | ccacgtatctatccagctcgtgcaactctaccttaaggaagcatggccagcaagtccaca | 2103 |
| | | | |
| Db | 2066 | ccactgcttattattttggcctggcgaacttgacactaaggagagtgtgccacagcttccag | 2125 |
| | | | |
| OY | 2104 | tgtccctctcttgctcttaaggagacagtgcccaagcact-----aggcagcatccaa | 2156 |
| | | | |
| Db | 2126 | cacttccctccctccgttcagggctcagagctgcccacgctttatcagagcagcagctccaagc | 2185 |
| | | | |
| OY | 2157 | ccaggtctaacgcccgaagcctttgacggctatcttccct--ggggcgttaatttgyg | 2214 |
| | | | |
| Db | 2186 | ccagagccagatgtccactcttttcgctgtgctcttcttcttgaggggcttatcatgtgtag | 2245 |
| | | | |
| OY | 2215 | atgaagcccttgagccaacaagggagacagcgatcatccacaagtcatgtgtctcatgcacctc- | 2273 |
| | | | |
| Db | 2246 | atpaaagccctgtagtgaggaagagcagtgagatccactgctatagcgtttgatpcaatcttca | 2305 |
| | | | |
| OY | 2274 | tcccttcccttccagcaacacttgtagtattgctcgtgcacatgtaacctgcgaaaaaga | 2329 |
| | | | |
| Db | 2306 | aacttttcccttccacgacacagagattttgtgcttggcatgcaacctgcgaaaaagaa | 2361 |
| | | | |

```

RESULT      8
AC025880
LOCUS       AK025880               1587 bp      mRNA      linear      PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ22227 fls, clone HRC01782, highly similar to
            HSA387747 Homo sapiens mRNA for stalin.
ACCESSION   AK025880
VERSION     AK025880
KEYWORDS    AK025880.1 GI:10438530
SOURCE      Oligo capping: fls (full insert sequence).
            Homo sapiens primary human renal epithelial cells cDNA to mRNA,
            clone_lib:HRC clone:HRC01782.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (stes)
AUTHORS     Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
            Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
            Nakamura,Y., Isogai,T. and Sugano,S.
            NEDO human cDNA sequencing project
            Unpublished (2000)
            2 (bases 1 to 1587)
            Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
            Direct Submission
            Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
            Sugano, Institute of Medical Science, University of Tokyo,
            Laboratory of Genome Structure Analysis, Human Genome Center;
            Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
            (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan: cDNA full insert
            sequencing: Research Association for Biotechnology: cDNA library
            construction, 5'- & 3'-end one pass sequencing: Department of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES
     source
     ..1..1587
     /organism="Homo sapiens"

```

| Query Match | Best Local Similarity | Score | DB | Length |
|----------------------------|---|---------------------------------------|----|--------|
| Matches 1027; Conservative | 80.2%; Pred. No. 5.1e-170; | 0; Mismatches 223; Indels 31; Gaps 7; | | |
| misc_feature | 1..1587 | /note="cloning vector pME185FL3" | | |
| | /note="highly similar to HSA387747 Homo sapiens mRNA for | | | |
| | stalin" | | | |
| BASE COUNT | 427 a 349 c 350 g 461 t | | | |
| ORIGIN | | | | |
| Query Match | 27.9%; Score 792.2; DB: 9; Length 1587; | | | |
| Best Local Similarity | 80.2%; Pred. No. 5.1e-170; | | | |
| Matches 1027; Conservative | 0; Mismatches 223; Indels 31; Gaps 7; | | | |
| 1060 | agaatgaggttttacctcagcagtcaccttattgagctgtgttatgatgatcctgtcgg | 1119 | | |
| 18 | acaaatgggtttttatcttcattgccttttttagcgtcttggtatgatgatccgtctg | 77 | | |
| 1120 | gtcaagctgtcagcaatttaagggcagaatggaatttccaactcgtggttcgaag | 1179 | | |
| 78 | gtcaagctgtcagcaatttaagggcagaatggaatttccaactcgtggttcgaag | 137 | | |
| 1180 | tttttagccttaagggatgttgagccttgatatttcgtgttgccgaggaattag | 1239 | | |
| 138 | tttttagccttaagggatgttgagccttgatatttcgtgttgccgaggaattag | 197 | | |
| 1240 | gtgtgatattcctctgtcgtgtgtatcctcctaacaatacaaaccttggagcctt | 1289 | | |
| 198 | gctgtgatattcctctgtcgtgtgtatcctcctaacaatacaaaccttggagcctt | 257 | | |
| 1300 | gctctcttgattagatcaacaacatctgagacattgctcctcgtatcgtatcctcc | 1359 | | |
| 258 | gctctcttgattagatcaacaacatctgagacattgctcctcgtatcgtatcctcc | 317 | | |
| 1360 | tgggcatacaaaatcactcttgccactatctctgtgaagtatgggccaatctgccaga | 1419 | | |
| 318 | tgggcatacaaaatcactcttgccactatctctgtgaagtatgggccaatctgccaga | 377 | | |
| 1420 | gtcttacccttgaagacactattggaagatgagcaacgtttcttgcacgcgtcgtga | 1479 | | |
| 378 | gtcttacccttgaagacactattggaagatgagcaacgtttcttgcacgcgtcgtga | 437 | | |
| 1480 | tcacatgatttggtagcaatcttcttccaacactatcgcgaagtggaatgcgaacatg | 1539 | | |
| 438 | tcacatgatttggtagcaatcttcttccaacactatcgcgaagtggaatgcgaacatg | 497 | | |
| 1540 | ccatgaatgataccaagaagacagaacatgaaggacccaataataatcctgtctcat | 1599 | | |
| 498 | ccatgaatgataccaagaagacagaacatgaaggacccaataataatcctgtctcat | 557 | | |
| 1600 | taattatccttggtagcaatcttcttccaacactatcgcgaagtggaatgcgaacatg | 1659 | | |
| 558 | taattatccttggtagcaatcttcttccaacactatcgcgaagtggaatgcgaacatg | 612 | | |
| 1660 | atcatatacaagaataaattgttactagaataaattgtttagattgtgaagcgttgaat | 1719 | | |
| 613 | atcatatacaagaataaattgttactagaataaattgtttagattgtgaagcgttgaat | 663 | | |
| 1720 | catgaatgataccaagaagacagaacatgaaggacccaataataatcctgtctcat | 1779 | | |
| 664 | catgaatgataccaagaagacagaacatgaaggacccaataataatcctgtctcat | 723 | | |
| 1780 | gtttcttggaacttcaatctcgaagtcacatatctcgtgtcgaatcgaagccaaccaat | 1839 | | |
| 724 | gtttcttggaacttcaatctcgaagtcacatatctcgtgtcgaatcgaagccaaccaat | 783 | | |
| 1840 | agggagatcttacttta-cttataagacacatacctaaagagatgagcgtgaataac | 1898 | | |
| 784 | agggagatcttacttta-cttataagacacatacctaaagagatgagcgtgaataac | 843 | | |
| 1898 | ctataccttgccttaataaaggatgaataatctcagctcgttgcgtgaataacatcgtt | 1958 | | |

| | | | |
|----|------|---|------|
| Db | 844 | TGATACCTTGGTTATTTAAACTACATGATAATTTCTCAGTACTGATTAACACCTGTCT | 903 |
| QY | 1959 | tgtaaaccttccctcaaaaaatatttgcatacgaacatccctgacatgagtgcttcaaac | 2018 |
| Db | 904 | TGTTACATTTCTCCTATAAAA-----ATTGTCAGCCCTCTCTGACACTTAGACCTTCAAC | 957 |
| QY | 2019 | tttagcctctcccaaggagctggcagcacatgtataatcaatcagcctggcaactcactgag | 2078 |
| Db | 958 | TTTAGCATCTCTGTGGACATGTGCATCCACTGTATTAATTTGGCTTGGCAACTGGACTGAGG | 1017 |
| QY | 2079 | gaagcatgcccagaggaagtgcacatgccccctctctctggtcttcaaggacagtgcccagca | 2138 |
| Db | 1018 | GGAGGTGCCCAAGCAGCGTGCACAACTCCCTCCCTG6CTTCAG6GTGACAGTCCCCAG | 1077 |
| QY | 2139 | ctt-----aggacagatcccaagaccaggggtgcagcgccaagctttgagcgtattctt | 2191 |
| Db | 1078 | CGTTTATAGAGGCGACGCTCCAGGCCCGAGGCCAATGTGTGACTCTTCCGCTGTGCTTT | 1137 |
| QY | 2192 | ccccct-9gggctgttaatgtgtgataagcccttgaaaccaacagaggaacagcgcgatcca | 2249 |
| Db | 1138 | CCTCGAGGGGCTATCATTGTGTAGATTAAGCCCTGAGTGAAGCAGAAGAGATGAGATCCA | 1197 |
| QY | 2250 | cagtcataggttccaatgcacccctc-tcccttcccttcccaagacacatgagatattgctg | 2308 |
| Db | 1198 | CTGTGATGTGCTGTGATACATCCTCAAACTTTCCCTTCCAGCACAAGAGAAATATTGGCTG | 1257 |
| QY | 2309 | gcatagtaacctgcaaaaagaaa | 2329 |
| Db | 1258 | GCATGCCAACCTGCCAAAAGAAA | 1278 |

RESULT 9
AK026921

Accession
AK026921

| | |
|----------|--|
| Source | Homo sapiens colon cDNA to mRNA, clone_11b:COL clone:COL08932. |
| Organism | Homo sapiens |

REFERENCE

AUTHORS

1 (sites)

Kawabata, A., Hikiiti, T., Kobayake, N., Inagaki, H., Ikema, Y.,
Kamada, S., Dachiwa, J., Imoto, S., Katsunuma, M., Nomura,
T.

TITLE
JOURNAL
SUGIYAMA, I., IANAKA, I., NAKAMURA, I., ISOGLA, I., and SUGANO, S.
NEDO human cDNA sequencing project
Unpublished (2000)

TITLE Direct Submission
AUTHORS Sugano, S., Suzuki, I., Uda, I., Ueda, S., M., Nishii, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.

Shirokane-dai 4-6-1 Minato-ku Tokyo 108-8630 Japan
Laboratory of Genome Structure Analysis, Human Genome Center,
Sugano, Institute of Medical Science, University of Tokyo,

COMMENT
Fax: 81-3-5449-5416)
NEDO human CDNA sequencing project supported by Ministry of
International Trade and Industry of Japan. cDNA full length

Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo, (partially supported by Science and Technology)

| FEATURES | location/Qualifiers |
|----------|---------------------|
| source | 1. .2006 |

```
/clone="COL08932"  
/clone_lib="COL"
```

| | | | |
|--------------|---------|------------------------------------|----------------------------------|
| misc_feature | 1..2006 | /issue="colon" | /note="cloning vector pME18SFL3" |
| misc_feature | 1..2006 | /note="highly similar to HSA387747 | stalin" |
| BASE COUNT | 559 a | 418 c | 420 g |
| ORIGIN | | | 609 t |

| | | | | |
|-----------------------|-----------------|-------------------|------------|--------------|
| Query Match | 21.2%; | Score 602.8; | DB 9; | Length 2066; |
| Best Local Similarity | 78.3%; | Pred. No. 8e-127; | | |
| Matches 822; | Conservative 0; | Mismatches 197; | Indels 31; | Gaps 7; |

Db 1 GAAGCTTTGCTCTTCGATTAGCATCCACCATCTGGATATGCTCCCTGATGCTG 60

Db 61 GATCCTCCCTGGGCAATCACAATATACATTGGCACTATTCCAGGAAATGGTTGGGCCCCGCA 120

D6 121 TTGCTAAAGTCGTGACCCCTGATAACACTGTTGGAGATGGCAACCGTGTCTATATG 180

Db 181 CCGCTGCTATTAATGTTTTGGGCCAATTTCTTACACTATTCGCCAAGGGAAGTAC 240

Db 241 AAACCTGGGCTCTCAATGATGACACCATGGACACAGACACTGAAGGAGACCAATAATAATCC 300

Db 301 TGCCTATTATGTATTTTATTATCATGTAACTCAAGGCCCT-----CTGTATT 355

[illegible]

Db 407 GCCTATATCATGAATATCTACTAGTTCGCCAGATAATAAATGACCTGCTTTAATAT 466

Db 467 GATTAATATGTAGCTAGGACCTCTACTTTTAAAGTTTCACATACCTGCGCTAATGGGCA 526

527 ACATGAGTAGGACAGTCTCTGTGATTTTTTATGGCCATACATAAAGGGAATGAGCTGAA 586
1001 accacacacacagggagag llllllllllll llllllllllll llllllllllll llllllllllll
1005 llllllllllll llllllllllll llllllllllll llllllllllll llllllllllll llllllllllll

587 CAGACGCGCTGATTACCTTTGCTTAAATTAAACAGATGATATTAATCTGAGGTACGATGAAC 646
1050 CAGACGCGCTGATTACCTTTGCTTAAATTAAACAGATGATATTAATCTGAGGTACGATGAAC 1743

[illegible][illegible][illegible][illegible]


```

QY      2250   cagtcataggttccatgcacacctc-tcccctccctcccaagcacactygaagtattgctcy 2308
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      43854   CCGCATATGCTTGTAGACATCCTCAAACTTTTCCTCCACGACAGAGAATAATTGGCTG 43913
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      2309   gcatgtaaccttcacaagaaga 2329
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db      43914   GCATCAACTCTGCAAAAGCAA 43934

RESULT 11
LOCUS    HSJ397H23              113202 bp     DNA             linear    PRI 20-MAR-2001
DEFINITION Human DNA sequence from clone RP3-397H23 on chromosome 6q12-14.1, complete sequence.
ACCESSION ALI21972
VERSION    ALI21972
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 113202)
AUTHORS   Kay,M.
TITLE      Direct Submission
COMMENT    Submitted (17-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
            Requests: clonerequests@sanger.ac.uk
            On Mar 12, 2001 this sequence version replaced gi:12832031..
            During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.; EMBL; SW.; SWISSPROT.; Tr.; TREMBL; Wp.; WormPep; Information on the WormPep database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep
            This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6
            RP3-397H23 is from the library RPrCl-3 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP3-397H23.

FEATURES
         Source
         location/Qualifiers
           1..113202
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="6"
             /map="q12-14.1"
             /clone="RP3-397H23"
             /clone_lib="RPrCl-3"
           1..58
             /note="AluU/F1AM repeat: matches 5..62 of consensus"
           260..570
             /note="Alusx repeat: matches 1..305 of consensus"
           674..707
             /note="MER20 repeat: matches 29..62 of consensus"
           708..880
             /note="AluSp repeat: matches 120..292 of consensus"
           961..1095

```

| | |
|---------------|--|
| repeat_region | /note="AluSg1 repeat: matches 1. .135 of consensus" 1105. .1428 |
| repeat_region | /note="AluSx repeat: matches 1. .306 of consensus" 1432. .1585 |
| repeat_region | /note="MER0 repeat: matches 60. .218 of consensus" 1826. .2123 |
| misc_feature | /note="Alu1 repeat: matches 1. .295 of consensus" complement(2112. .2118) |
| misc_feature | /note="Single clone region. assembly confirmed by restriction digest data" complement(2120. .2127) |
| misc_feature | /note="Single clone region. assembly confirmed by restriction digest data" complement(2129. .2147) |
| misc_feature | /note="Single clone region. assembly confirmed by restriction digest data" complement(2149. .2160) |
| repeat_region | /note="Single clone region. assembly confirmed by restriction digest data" 2150. .2458 |
| misc_feature | /note="AluSx repeat: matches 1. .306 of consensus" complement(2161. .2164) |
| misc_feature | /note="Single clone region. assembly confirmed by restriction digest data" complement(2168. .2263) |
| misc_feature | /note="Single clone region. assembly confirmed by restriction digest data" 2468. .2774 |
| repeat_region | /note="AluSg1 repeat: matches 1. .308 of consensus" 2795. .2932 |
| repeat_region | /note="FLAM_C repeat: matches 1. .118 of consensus" 3028. .3325 |
| repeat_region | /note="Alu1 repeat: matches 1. .296 of consensus" 3355. .3644 |
| repeat_region | /note="AluSg repeat: matches 1. .287 of consensus" 3775. .4074 |
| repeat_region | /note="AluSx repeat: matches 1. .300 of consensus" 4096. .4386 |
| repeat_region | /note="Alu10 repeat: matches 1. .298 of consensus" 4490. .4793 |
| repeat_region | /note="Alu1b repeat: matches 1. .306 of consensus" 5117. .5253 |
| repeat_region | /note="FLAM_C repeat: matches 1. .132 of consensus" 5501. .5814 |
| repeat_region | /note="AluSg repeat: matches 1. .308 of consensus" 6097. .6225 |
| repeat_region | /note="FLAM_A repeat: matches 1. .131 of consensus" 6296. .6479 |
| repeat_region | /note="Alu1b repeat: matches 129. .312 of consensus" 6545. .6830 |
| repeat_region | /note="Alu10 repeat: matches 42. .306 of consensus" 6965. .7270 |
| repeat_region | /note="AluSg repeat: matches 1. .306 of consensus" 7308. .7602 |
| repeat_region | /note="Alu1b repeat: matches 1. .295 of consensus" 7737. .8035 |
| repeat_region | /note="AluSg repeat: matches 1. .299 of consensus" 8184. .8488 |
| repeat_region | /note="AluSx repeat: matches 1. .299 of consensus" 8917. .9176 |
| repeat_region | /note="AluSg1 repeat: matches 1. .293 of consensus" 9212. .9377 |
| repeat_region | /note="L1M4 repeat: matches 6005. .6177 of consensus" 9490. .9785 |
| repeat_region | /note="AluSx repeat: matches 1. .296 of consensus" 10352. .10431 |
| repeat_region | /note="L2 repeat: matches 2616. .2702 of consensus" 10432. .10734 |
| repeat_region | /note="AluSx repeat: matches 1. .301 of consensus" 10735. .10778 |
| repeat_region | /note="L2 repeat: matches 2575. .2616 of consensus" 10846. .10985 |
| repeat_region | /note="MIR repeat: matches 53. .200 of consensus" |

```

repeat_region      11023..11140
/note="FLAM_C repeat: matches 1..118 of consensus"
repeat_region      12005..12079
/note="L1MD3 repeat: matches 7662..7737 of consensus"
repeat_region      12350..12647
/note="AlusX repeat: matches 1..297 of consensus"
repeat_region      12669..12916
/note="AlusX repeat: matches 1..293 of consensus"
repeat_region      13058..13351
/note="AlusC repeat: matches 1..296 of consensus"
repeat_region      13504..13657
/note="AluY repeat: matches 152..306 of consensus"
repeat_region      13888..13947
/note="AlusX repeat: matches 25..138 of consensus"
repeat_region      13948..14254
/note="Alusg repeat: matches 1..305 of consensus"
repeat_region      14235..14437
/note="AlusX repeat: matches 138..307 of consensus"
repeat_region      14482..14610
/note="MERB repeat: matches 46..167 of consensus"
repeat_region      14614..14911
/note="AlusC repeat: matches 1..301 of consensus"
repeat_region      14934..15229
/note="Alusg repeat: matches 1..297 of consensus"
repeat_region      15266..15556
/note="Aluo repeat: matches 5..299 of consensus"
repeat_region      15578..15888
/note="Aluya5 repeat: matches 1..311 of consensus"
repeat_region      15965..16082
/note="FLAM_C repeat: matches 1..118 of consensus"
repeat_region      16157..16439
/note="Alub repeat: matches 11..294 of consensus"
repeat_region      16518..16913
/note="Alusg repeat: matches 1..296 of consensus"
repeat_region      16933..16998
/note="33 copies 2 mer tt 69% conserved"
repeat_region      17001..17104
/note="Alusg/x repeat: matches 196..307 of consensus"
repeat_region      17135..17428
/note="Alusg repeat: matches 1..293 of consensus"
repeat_region      17790..18090
/note="AlusX repeat: matches 1..301 of consensus"
repeat_region      19220..19547
/note="Alub repeat: matches 2..290 of consensus"
repeat_region      20724..21025
/note="Aluo repeat: matches 1..304 of consensus"
repeat_region      21045..21082
/note="19 copies 2 mer gt 89% conserved"
repeat_region      21084..21260
/note="Alusg/x repeat: matches 134..310 of consensus"
repeat_region      21367..21670
/note="AlusX repeat: matches 1..306 of consensus"
repeat_region      21819..21939
/note="MIR repeat: matches 32..165 of consensus"
repeat_region      22647..22843
/note="AlusX repeat: matches 1..299 of consensus"
repeat_region      23011..23316
/note="AluY repeat: matches 1..306 of consensus"
repeat_region      23399..23702
/note="Alusg repeat: matches 1..304 of consensus"
repeat_region      24332..24622
/note="AlusX repeat: matches 1..302 of consensus"
repeat_region      24680..24971
/note="AlusC repeat: matches 1..292 of consensus"
repeat_region      25178..25302
/note="Alusg/x repeat: matches 172..300 of consensus"

Query Match      16.9%: Score 480.2; DB 9; Length 113202;
Best Local Similarity 75.7%: Pred. No. 1.2e-98;
Matches 697; Conservative 0; Mismatches 193; Indels 31; Gaps 7.

```

| | | | | |
|------------|--|-----------|--|--------|
| D | b | 56637 | gctcttcctttatagaacactgttggaagatggcacaacctgtttcatatattgctgctgca | 56578 |
| Q | y | 1480 | lcaatgtaattggygccattttcttcacactlccgccaagtygaatgycanaacctgg | 1539 |
| D | b | 56577 | tttaaatgttttggtgccattttctttacacatgttcgccaagaagtgaaactggg | 56518 |
| Q | y | 1540 | ccatgaatgataccaagaagacagaactctgaagaacaaataataatccgtctcat | 1599 |
| D | b | 56517 | ctctcaattgataccatggacacacacactgaaagaaaccaaataatcctgcctcat | 56458 |
| Q | y | 1600 | taatgatcttgattcatatgtaacctaaagtygccttgaattttaatgtygaagca | 1659 |
| D | b | 56457 | taatgatattttttatttatcatcttaacctcaaaagtgccctt----ctgatttctgaaaga | 56403 |
| Q | y | 1660 | atctatatacaagataaanaatgtaactagaanaaatgtygttagatttgaagctgta | 1719 |
| D | b | 56402 | ttcttatgt-----cttttttttaattgtactgtattagatttttaagccctatatt | 56352 |
| Q | y | 1720 | catgaatgtaactatggttgcacataaagcaaaatagactatttttaattataacc | 1779 |
| D | b | 56351 | catgaanaatgtaactggttgcacataaagcaaaatagactatttttaattataacc | 56292 |
| Q | y | 1780 | gtttctgtaaatcaaatcaatcagggtacatatctcggctgcaagtcagaacccacat | 1839 |
| D | b | 56291 | gttaactgtaaatcaaatcaatcagggtacatatctcggctgcaagtcagaacccacat | 56232 |
| Q | y | 1840 | aggggagttctatcta-ttataagaacatacctaagaatgagctgaataagaccct | 1898 |
| D | b | 56231 | aggagactgtctgttatatttttttagggccaaacttaaaaggaattgactgaaacagactcc | 56172 |
| Q | y | 1899 | ctataccttctgtaataaagtygaataataatctcaagctctgtaataacatcgctt | 1958 |
| D | b | 56171 | tgatgactgttggcttaattaaactgaatgatatttctcaggtactatataacacctgtgt | 56112 |
| Q | y | 1959 | tgtaaccttctcaaaaaaattattgtcatcagaatccctgacatgtagtctgaac | 2018 |
| D | b | 56111 | tttttacttttctctctatpaaaa-----attgtcagctctctctgacactttagacctcaaac | 56058 |
| Q | y | 2019 | tttaagcttctcagaagagctggaacacactgtatattcaatcagctctgcaacttcaatg | 2078 |
| D | b | 56057 | ttttagactctgtgtgagagctgccatccactgtaattttcgcttgcaaacctgagctgag | 55998 |
| Q | y | 2079 | gaagcatgcccagaagctgacatgtccctctctggtcttaaggaagtgccagca | 2138 |
| D | b | 55997 | ggagtggtgccacagcagcagctccactccctccctgacagggctcagagtcaccag | 55938 |
| Q | y | 2139 | ctt-----aggagacatcaagaagcagagtgctcagcgaagcttggacgttctt | 2191 |
| D | b | 55937 | cttttatatcagagagcagctccacagccacagcagctgctgactcttccgctgctctt | 55878 |
| Q | y | 2192 | ccctt--ggagctgttaatgtgtgaatgaagccttgaagccacagagaacagcgcgatcca | 2249 |
| D | b | 55877 | ccttgagagggcgtatcattatgtgtatgaatgaagccctgagatgacacagacatgacatcca | 55818 |
| Q | y | 2250 | caatcatggttccatgcaacctc--tcccttcccttcccagcaacatgagatattgctg | 2308 |
| D | b | 55817 | ctgcatatgctgtttgattatcatctctcaaaactttcccttccagcacagagaattatggctg | 55758 |
| Q | y | 2309 | gcatgtaacctgcaaaagaa | 2329 |
| D | b | 55757 | gcattgaacctgcaaaagaa | 55737 |
| RESULT 12 | | | | |
| AC034271 | | | | |
| LOCUS | AC034271 | 149597 bp | DNA | linear |
| DEFINITION | Homo sapiens chromosome 6 clone RP11-598G19 map 6, WORKING DRAFT | | | |
| SEQUENCE | Sequence, 31 unordered pieces. | | | |
| ACCESSION | AC034271 | | | |
| VERSION | AC034271.2 GI:7770480 | | | |
| KEYWORDS | HTG; HTGS_PHASE1; HTGS_DRAFT. | | | |
| SOURCE | human. | | | |
| ORGANISM | Homo sapiens | | | |

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 149597)
Birren B., Linton L., Nusbaum C. and Lander E.
Homo sapiens chromosome 6, clone RP11-598G19
Unpublished
2 (bases 1 to 149597)
Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,
Anderson S., Baldwin J., Barna N., Bastien V., Beld F.,
Boguslavsky L., Boukhgalter B., Brown A., Burkett G.,
Campopiano A., Castle A., Choepel I., Colangelo M., Collins S.,
Collumore A., Cooke P., DeArillano K., Dewar K., Diaz J.S.,
Dodge S., Domino M., Doyle M., Ferrera P., Fitzhugh W., Gage D.,
Galagan J., Gardyna S., Glade S., Goyette M., Graham L.,
Grand-Pierre N., Grant G., Hagos B., Heatford A., Horton L.,
Howard J.C., Iliev I., Johnson R., Jones C., Kann L., Karatas A.,
Klein J., Lacombe K., Lamazares R., Landers T., Lehotzky J.,
Levine R., Lieu C., Liu G., Locke K., MacDonald P., Margulis N.,
McCarthy M., McEwan P., McGurk A., McKernan K., McPheters R.,
Melidrim J., Menus L., Mihova T., Miranda C., Mienga V., Morrow J.,
Murphy T., Naylor J., Norman C.H., O'Connor T., O'Donnell P.,
O'Neill D., Oliver T.M., Oliver J., Peterson K., Pierre N.,
Pisani C., Pollara V., Raymond C., Riley R., Rogov P., Rothman D.,
Roy A., Santos R., Schauer S., Severy P., Spencer B.,
Stange-Thomann N., Stojanovic N., Subramanian A., Talamas J.,
Testaye S., Theodore J., Tirrell A., Travers M., Trigilio J.,
Vassiliev H., Viel R., Vo A., Wilson B., Wu X., Wyman D., Ye W.J.,
Young G., Zainoun J., Zimmer A. and Zody M.

TITLE
JOURNAL
COMMENT

Submitted (05-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 12, 2000 this sequence version replaced gi:7417823.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L7098
Center clone name: 598_G.19

Summary Statistics
Sequencing vector: M13: M77815, 100% of reads

Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 134701 bases at least Q40
Consensus quality: 141812 bases at least Q40
Consensus quality: 144823 bases at least Q20

Insert size: 166000; agarose-fp
Insert size: 146597; sum-of-coverage
Quality coverage: 3.6 in Q20 bases; agarose-fp

Quality coverage: 4.1 in Q20 bases; sum-of-coverage

* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1019 1118: contig of 1018 bp in length
* 1019 1118: gap of 100 bp
* 1119 2203: contig of 1085 bp in length
* 2204 2303: gap of 100 bp
* 2304 3814: contig of 1511 bp in length
* 3815 3914: gap of 100 bp
* 3915 4327: contig of 413 bp in length
* 4328 4427: gap of 100 bp
* 4428 6447: contig of 2020 bp in length
* 6448 6547: gap of 100 bp

FEATURES

| source | FEATURES |
|--------------|--|
| misc_feature | 6548 8884: contig of 2337 bp in length |
| misc_feature | 8885 8984: gap of 100 bp |
| misc_feature | 8985 10630: contig of 1646 bp in length |
| misc_feature | 10631 10730: gap of 100 bp |
| misc_feature | 10731 13202: contig of 2472 bp in length |
| misc_feature | 13203 13302: gap of 100 bp |
| misc_feature | 13303 15992: contig of 2690 bp in length |
| misc_feature | 15993 16092: gap of 100 bp |
| misc_feature | 16093 18722: contig of 2630 bp in length |
| misc_feature | 18723 18822: gap of 100 bp |
| misc_feature | 18823 21695: contig of 2873 bp in length |
| misc_feature | 21696 21795: gap of 100 bp |
| misc_feature | 21796 25265: contig of 3470 bp in length |
| misc_feature | 25266 25365: gap of 100 bp |
| misc_feature | 25366 28014: contig of 2649 bp in length |
| misc_feature | 28015 28114: gap of 100 bp |
| misc_feature | 28115 31378: contig of 3264 bp in length |
| misc_feature | 31379 31478: gap of 100 bp |
| misc_feature | 31479 34284: contig of 2806 bp in length |
| misc_feature | 34285 34384: gap of 100 bp |
| misc_feature | 34385 37707: contig of 3323 bp in length |
| misc_feature | 37708 37807: gap of 100 bp |
| misc_feature | 37808 41922: contig of 4115 bp in length |
| misc_feature | 41923 42022: gap of 100 bp |
| misc_feature | 42023 45878: contig of 3856 bp in length |
| misc_feature | 45879 45978: gap of 100 bp |
| misc_feature | 45979 49890: contig of 3912 bp in length |
| misc_feature | 49891 49990: gap of 100 bp |
| misc_feature | 49991 54041: contig of 4051 bp in length |
| misc_feature | 54042 54141: gap of 100 bp |
| misc_feature | 54142 59531: contig of 5390 bp in length |
| misc_feature | 59532 59631: gap of 100 bp |
| misc_feature | 59632 65537: contig of 5966 bp in length |
| misc_feature | 65538 65637: gap of 100 bp |
| misc_feature | 65638 71486: contig of 5849 bp in length |
| misc_feature | 71487 71586: gap of 100 bp |
| misc_feature | 71587 76694: contig of 5108 bp in length |
| misc_feature | 76695 76794: gap of 100 bp |
| misc_feature | 76795 82579: contig of 5785 bp in length |
| misc_feature | 82580 82679: gap of 100 bp |
| misc_feature | 82680 90920: contig of 8241 bp in length |
| misc_feature | 90921 91020: gap of 100 bp |
| misc_feature | 91021 99034: contig of 8014 bp in length |
| misc_feature | 99035 99134: gap of 100 bp |
| misc_feature | 99135 110324: contig of 11190 bp in length |
| misc_feature | 110325 110424: gap of 100 bp |
| misc_feature | 110425 119810: contig of 9386 bp in length |
| misc_feature | 119811 119910: gap of 100 bp |
| misc_feature | 119911 130142: contig of 10332 bp in length |
| misc_feature | 130143 130242: gap of 100 bp |
| misc_feature | 130243 149597: contig of 19555 bp in length. |

Location/Qualifiers

1. 149597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-598G19"
/clone_lib="RPC1-11 Human Male BAC"
1. 1018
/note="assembly-fragment"
1119. 2203
/note="assembly-fragment"
2304. 3814
/note="assembly-fragment"
3915. 4327
/note="assembly-fragment"
clone_end:17
vector_side:right"
4428. 6447
/note="assembly-fragment"
6548. 8884
/note="assembly-fragment"

```
misc_feature      8985..10630
                  /note="assembly_fragment"
misc_feature      10731..13202
                  /note="assembly_fragment"
misc_feature      13303..15992
                  /note="assembly_fragment"
misc_feature      16093..18722
                  /note="assembly_fragment"
misc_feature      18823..21695
                  /note="assembly_fragment"
misc_feature      21796..25265
                  /note="assembly_fragment"
misc_feature      25366..28014
                  /note="assembly_fragment"
misc_feature      28115..31378
                  /note="assembly_fragment"
misc_feature      31479..34284
                  /note="assembly_fragment"
misc_feature      34385..37707
                  /note="assembly_fragment"
misc_feature      37808..41922
                  /note="assembly_fragment"
misc_feature      42023..45878
                  /note="assembly_fragment"
misc_feature      45979..49890
                  /note="assembly_fragment"
misc_feature      49991..54041
                  /note="assembly_fragment"
misc_feature      54142..59531
                  /note="assembly_fragment"
misc_feature      59632..65537
                  /note="assembly_fragment"
misc_feature      65638..71486
                  /note="assembly_fragment"
misc_feature      71587..76694
                  /note="assembly_fragment"
misc_feature      76795..82579
                  /note="assembly_fragment"
misc_feature      82680..90920
                  /note="assembly_fragment"
misc_feature      91021..99034
                  /note="assembly_fragment"
misc_feature      99135..110324
                  /note="assembly_fragment"
misc_feature      110425..119810
                  /note="assembly_fragment"
```

```
Query Match      16.9%; Score 480.2; DB 2; Length 149597;
Best Local Similarity 75.7%; Pred. No. 1.2e-98;
Matches 697; Conservative 0; Mismatches 193; Indels 31; Gaps 7;
```

```
QY 1420 gcttaccctgagacacatctgagaaatgagcaactgtttctgcatcgctgctcta 1479
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75367 GTCTTCCTTTATGACAACTGTTGGAGAAATGGCAAAACCGTCTTATATGCTGCTGCTA 75426

QY 1480 tcaatgatactgtgcatctttcttcacacatctgcgaagtgtaagtgcaaaactggg 1539
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75427 TTATATGTTTTGGTGCCATTTCTTTACATATTCGCCAAAGGGAAGTACAAAACTGGG 75486

QY 1540 ccatcagtgatcaccaagagacacagaatggaagaaacataataatcctgtctctat 1599
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75487 CTCCTCAATGATGACCAAGGACACAGACACGAAAGAACCAATAAATATCTCGCTCTAT 75546

QY 1600 taaatgatactgttatacatgtaacctaagagccttgatataatttaatggttaagca 1659
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75547 TAAATGATTTTTTATTTTATCATGTAACCTCAAGTGCCCTT-----CTGTATTTGTTAAGCA 75601

QY 1660 atctataacagaataaattgtaactagaataattggttagattgtlaagggctgtaat 1719
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75602 TTCTATATG-----CTTTTATTAATGTGACTGTGATTAAGATTTTAAAGCCTATATAT 75652

QY 1720 catgaatgtaactagtgcatataagcaaaattagctattttaattatttaacc 1779
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 75653 CATGAATATCATAGTACTAGTTGCCAGAAATATAATGAACGTGTGTTAATATGATAATAT 75712
QY 1780 gtttcggaacttacatcatcaggtacacatctgctcgaatgcaagcaaccacaat 1839
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75713 GTAAGCTAGAGACTTCTACTTTTGGTTTCACATACCTGCTCTGTATGCGGCAACATGAACT 75772

QY 1840 agggagttctatla-ttataagacatacctaaagatgagctgaaatagaccct 1898
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75773 AGGACACTTCTGTGATTTTAAAGGCGCATACATAAAGGAATAGCTGAAGAACAGACTCC 75832

QY 1899 ctataccttgcttaataagtgagataaatctcaggtcttgtaaacatctgttt 1958
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75833 TGAATACCTTGCTTAATTAACATAGATGAATATTCAGGTACTGATTAACACCTGTTGT 75892

QY 1959 tgaacacttcctcaaaaattatttgtaacgaatcccttcacatgtagtgcctaacc 2018
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75893 TGTTCACCTTCTCATATAA-----ATTGTACAGCTCTCTGTACACTTAGACCTCAAC 75946

QY 2019 tttagcctcccaagagctgagcagccactgatacattcagcctgagcaacttaactagag 2078
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75947 TTTAGACATCTCTGTGGAGCTGCGCATCCAGTATTAATTTGCGCTGGCACTGAGCTGAGG 76006

QY 2079 gaagcatgccagagcagctgacacatgtccctctctgcttcaaggagcagtgccagca 2138
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 76007 GGAAGTGCCCCAGACAGCTGCAAGCACTCCCTCGCTGAGGCTCAGAGTCCAG 76066

QY 2139 ctt-----aggcagatccaagacagaggttagcagcgaagctttagacgttatctt 2191
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 76067 CGTTTATCAAGAGCAGCATCCAAAGCCAGAGCCAGTGTACATCTTGGCTGGCTTT 76126

QY 2192 cccct--ggggctgttaatgtgtgtaagccctgagcccaagagcagcagatca 2249
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 76127 CCTGTGAGGGGATCATCAATGTGATGAATGAAGCCCTGAGTGAAGAGCAGTGAATCA 76186

QY 2250 caatgcatggttcacatgcacccctc-tccctccctcccaagcagcaggtattgctg 2308
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 76187 CTGTATGTTGTGATTCATCTCAAACTTTCCCTTCCAGCACAGAGAAATATGTGCTG 76246

QY 2309 gcaatgtaactgcaaaagaa 2329
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 76247 GCATGCAACTGCMAAAGAAA 76267
```

```
RESULT 13
AC025535 157749 bp DNA linear HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 6 clone RP11-206H23, WORKING DRAFT
DEFINITION
SEQUENCE 25 unordered pieces.
ACCESSION AC025535
VERSION AC025535.4 GI:7658462
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
```

```
REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 157749)
JOURNAL The sequence of Homo sapiens clone
REFERENCE 2 (bases 1 to 157749)
AUTHORS Waterston,R.H.
JOURNAL Submitted (10-MAR-2000) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
```

On/Apr 28, 2000 this sequence version replaced gi:7582731.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0206H23


```

----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-terminator ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 14137 bases at least Q40
Consensus quality: 14682 bases at least Q30
Consensus quality: 14842 bases at least Q20
Insert size: 16800; agarose-fp
Insert size: 155349; sum-of-contigs
Quality coverage: 3.52 in Q20 bases; agarose-fp
Quality coverage: 3.92 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1470: contig of 1470 bp in length
1471 1570: gap of unknown length
1571 3019: contig of 1449 bp in length
3020 3119: gap of unknown length
3120 5356: contig of 2237 bp in length
5357 5457: gap of unknown length
5457 7747: contig of 2291 bp in length
7748 7848: gap of unknown length
7848 9214: contig of 1366 bp in length
9214 9314: gap of unknown length
9314 10753: contig of 1440 bp in length
10754 10854: gap of unknown length
10854 12243: contig of 1390 bp in length
12244 12343: gap of unknown length
12344 14551: contig of 2208 bp in length
14552 14652: gap of unknown length
14652 17066: contig of 2415 bp in length
17067 17167: gap of unknown length
17167 19602: contig of 2436 bp in length
19603 19702: gap of unknown length
19703 23843: contig of 4141 bp in length
23844 23943: gap of unknown length
23944 27783: contig of 3840 bp in length
27784 27883: gap of unknown length
27884 31792: contig of 3809 bp in length
31793 31892: gap of unknown length
31893 35687: contig of 3795 bp in length
35688 35787: gap of unknown length
35788 39711: contig of 3924 bp in length
39712 39811: gap of unknown length
39812 44204: contig of 4393 bp in length
44205 44304: gap of unknown length
44305 48833: contig of 4529 bp in length
48834 48933: gap of unknown length
48934 56617: contig of 7684 bp in length
56618 56717: gap of unknown length
56718 63063: contig of 6346 bp in length
63064 63163: gap of unknown length
63164 74288: contig of 11125 bp in length
74289 74388: gap of unknown length
74389 87403: contig of 13015 bp in length
87404 87504: gap of unknown length
87504 99698: contig of 12195 bp in length
99699 99798: gap of unknown length
99799 115384: contig of 15586 bp in length
115385 115484: gap of unknown length
115485 132231: contig of 16647 bp in length
132232 132331: gap of unknown length
132332 157749: contig of 25518 bp in length.

* Location/Qualifiers
1. 157749

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-206H23"
BASE COUNT 43129 a 34263 c 34620 g 43330 t 2407 others
ORIGIN

Query Match 16.9%; Score 480.2; DB 2; Length 157749;
Best Local Similarity 75.7%; Pred. No. 1.2e-98;
Matches 697; Conservative 0; Mismatches 193; Indels 31; Gaps 7;

QY 1420 gttcacccttgagacacactatggaatgagcaactgtttctgcacgtcgtgta 1479
|||||
DB 93294 GTCCTCTTTATAGAACACTGTGGAGAAATGGCAACCGTGTCTATATTTGCTGCTGA 93295

QY 1480 tcaatgtttgtggtcattttcttcacacatttcgcaaaagtgaaagtgcaaaactg 1539
|||||
DB 93234 TTAATGTTTGGTCCATTTCTTTACACTATTCGCCAAAGGTGAACTACAAAACCTGG 93175

QY 1540 ccatcagtcacacacaaagacagaactgaagaacacaaatcctcctctctat 1599
|||||
DB 93174 CTCCTAATGATCCATGACACAGACACTGAGAACCAATATATTCCTGCTCTAT 93115

QY 1600 taatgtatcttgtttatcatgtaacttaaaagtccttgcataatttaatgtaagca 1659
|||||
DB 93114 TAATGATTTTATTTATCATGATACCTCAAAAGTCCCTT-----CTGATGTGTAGCA 93060

QY 1660 atctatacaaaagaataatgtaactagaataatgtgttagatttgaagctgtat 1719
|||||
DB 93059 TTCTATGT-----CTTTTAAATTTGTCTGTATATTAATTTTAAAGCCATATAT 93009

QY 1720 catgaatgtcactagttgcacataagaacaaataggcatttataattataaacc 1779
|||||
DB 93008 CATGAATATCTCTAGTGTCCGAAATTAATTAATTAATTAATTAATTAATTAAT 92949

QY 1780 gtttcgtgaacttcaattcagggtlcaacatactgcgtcgaagcaagcaaccacat 1839
|||||
DB 92948 GTAACCTAGACTTCTACTTTTGTGTTCAATACATACCTGCTGCTGAGCAACATGAGT 92889

QY 1840 agggaggtcttattta-tttataagacacatacctaagaagatgagctgaataagacctt 1898
|||||
DB 92888 AGGACAGTTCGTGTTATTTTAAAGGCAATCAAAAGGAAAGACTTAAACAGACTCC 92829

QY 1899 ctatcccttcttaataaagtgagtaataatctcaggtctgttaaacctcgtttt 1958
|||||
DB 92878 TGATACCTTTGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 92769

QY 1959 tgtacaccttctcaaaaatatttgcatacgaacatccctgacatgtagtctcaaac 2018
|||||
DB 92768 TGTTCACCTTCTCTCTATAAA-----ATTGTCAGCTCTCTGACACTTAAGCCTCAAC 92715

QY 2019 tttagcctctccacagagctggcagcaactgatalcctcagccttgcaactcaggagg 2078
|||||
DB 92714 TTTACATCTCTGTGGACCTGCATCTGATATTTTCGCTGCACTTGAAGTGAAGTGAAG 92655

QY 2079 gaagcaatcccaagagagctgcacacatgcccctcctcgtcttcaagggacagtcccaaca 2138
|||||
DB 92654 GGAAGTGTCCCAAGGAGCTGCCAAGCACTCCCTCTCTGCTTCAAGGTCAGGTGCCAG 92595

QY 2139 ctt-----aggcagcagcagcaagcagggctcagcgcagagccttggacgttactt 2191
|||||
DB 92594 CGTTATATCAGAGGAGGATCCAGCCAGAGCCAGGACATGTGCACTTTGGCGTGTGCCCTTT 92535

QY 2192 cccct--gggctgttaattgtgtgtaagaagccttgagcacaagagcagcagcgtatcca 2249
|||||
DB 92534 CCTCAGAGGGCTATCAATGTGTATGATTAAGCCCTGAGTGAAGCAGACAGTGAATCA 92475

QY 2250 cagtaatgttccatgacacctc-tcccctccctcccaagcaactgagatgctgctg 2308
|||||
DB 92474 CTGCTATGTCTTGTATATCTCAAACTTTCCTCCAGCAGCAGAGGAATATTTGGCTG 92415

QY 2309 gcatgtaacctgcaaaagaa 2329

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 13:52:14 ; Search time 592.65 Seconds
(without alignments)
8239.103 Million cell updates/sec

Title: US-09-776-865-3
Perfect score: 2844
Sequence: 1 cccggggggcgggggcttcg.....ataaaaaaaaaaaaaa 2844

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_032802:*
1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | ID | Description |
|------------|--------|---------|--------------|----|-------------|
| 1 | 2844 | 100.0 | 2844 | 21 | AAZ50876 |
| 2 | 2844 | 100.0 | 2844 | 22 | AAD10326 |
| 3 | 1572.4 | 55.3 | 2512 | 22 | AAF55900 |
| 4 | 1570.8 | 55.2 | 2930 | 21 | AAZ50879 |
| 5 | 1570.8 | 55.2 | 2930 | 22 | AAD10325 |
| 6 | 1567.6 | 55.1 | 2602 | 21 | AAZ50875 |
| 7 | 1458.8 | 51.3 | 2712 | 22 | AAK94876 |
| 8 | 1413 | 49.7 | 1485 | 21 | AAZ50880 |
| 9 | 1305 | 45.9 | 1485 | 21 | AAZ50881 |

| | | | | | |
|----|--------|------|-------|----|----------|
| 10 | 1200 | 42.2 | 1488 | 22 | AAI58115 |
| 11 | 1169.8 | 41.1 | 2670 | 22 | AAH79234 |
| 12 | 672.8 | 23.7 | 1975 | 22 | AAH99626 |
| 13 | 642.4 | 22.6 | 929 | 22 | AAI59901 |
| 14 | 594.8 | 20.9 | 853 | 22 | AAK93901 |
| 15 | 553 | 19.4 | 838 | 22 | AAK92364 |
| 16 | 456.2 | 16.0 | 1975 | 22 | AAH99626 |
| 17 | 341.8 | 12.0 | 798 | 23 | AAH56219 |
| 18 | 341.8 | 12.0 | 1066 | 23 | AAH52158 |
| 19 | 314.2 | 11.0 | 375 | 22 | AAH52158 |
| 20 | 229.8 | 8.1 | 349 | 22 | AAI13566 |
| 21 | 226.6 | 8.0 | 1641 | 23 | ABL18113 |
| 22 | 226.6 | 8.0 | 1786 | 23 | ABL07417 |
| 23 | 223.8 | 7.9 | 3422 | 22 | AAK52406 |
| 24 | 199 | 7.0 | 264 | 22 | AAI22435 |
| 25 | 192.6 | 6.8 | 2291 | 23 | ABL20393 |
| 26 | 192 | 6.8 | 1880 | 23 | ABL1735 |
| 27 | 190.6 | 6.7 | 1620 | 23 | ABL14775 |
| 28 | 190.6 | 6.7 | 3620 | 23 | ABL14774 |
| 29 | 190.6 | 6.7 | 90104 | 23 | ABL12402 |
| 30 | 186.8 | 6.6 | 2889 | 22 | AAI93801 |
| 31 | 185.2 | 6.5 | 2716 | 17 | AAI42064 |
| 32 | 185.2 | 6.5 | 2716 | 19 | AAV35303 |
| 33 | 164.2 | 5.8 | 2281 | 19 | AAV57909 |
| 34 | 160 | 5.6 | 1943 | 23 | ABL09181 |
| 35 | 159.8 | 5.6 | 1562 | 21 | AAAC4267 |
| 36 | 155.6 | 5.5 | 1643 | 19 | AAV43711 |
| 37 | 154 | 5.4 | 1795 | 19 | AAV57910 |
| 38 | 146 | 5.1 | 300 | 20 | AAZ14190 |
| 39 | 142.2 | 5.0 | 1939 | 23 | ABL03769 |
| 40 | 140.4 | 4.9 | 4899 | 23 | ABL03768 |
| 41 | 133 | 4.7 | 1512 | 23 | ABL15787 |
| 42 | 132.2 | 4.6 | 2003 | 23 | ABL08217 |
| 43 | 129.4 | 4.5 | 2280 | 23 | ABL05527 |
| 44 | 126.6 | 4.5 | 1440 | 23 | ABL12049 |
| 45 | 126.6 | 4.5 | 3440 | 23 | ABL12048 |

ALIGNMENTS

| | |
|----------|---|
| RESULT 1 | |
| AAZ50876 | AAZ50876 standard; cDNA; 2844 BP. |
| XX | XX |
| AC | AAZ50876; |
| XX | XX |
| DT | 31-MAY-2000 (first entry) |
| XX | XX |
| DE | Sheep GBS toxin receptor (SP55) cDNA. |
| XX | XX |
| KW | Sheep GBS toxin receptor; group B beta-haemolytic streptococci; SP55; |
| KW | pathological vascularisation; cancer metastases; anglogenesis; |
| KW | neovascularisation; reperfusion injury; scarring; keloid; |
| KW | chronic inflammation; rheumatoid arthritis; psoriasis; neural injury; |
| KW | endothelial cell proliferation; antibacterial; anticancer; |
| XX | anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss. |
| OS | Ovis sp. |
| XX | XX |
| FT | Key |
| FT | Location/Qualifiers |
| FT | CDS 84..1571 |
| FT | /*tag= a |
| FT | /product= "Sheep GBS toxin receptor" |
| XX | XX |
| PN | WO200005375-A1. |
| XX | XX |
| PD | 03-FEB-2000. |
| XX | XX |
| PF | 22-JUL-1999; 99WO-US16676. |
| XX | XX |
| PR | 22-JUL-1998; 98US-0093843. |
| XX | XX |

PA (UYVA-) UNIV VANDERBILT.
XX
PI HELLERGVIST CG, FU C;
XX
DR WPI: 2000-205377/18.
DR P-PSDB: AA145086.
XX
PT New polynucleotide encoding mammalian receptor for streptococcus toxin,
PT useful for diagnosis and treatment of, e.g. pneumonia in neonates -
XX
PS Claim 3; Page 83-86; 109pp; English.
XX
CC The present cDNA sequence encodes partial sheep GBS (group B beta
CC -haemolytic streptococci) toxin receptor (SP55). This sequence was cloned
CC using a primary culture of sheep lung endothelial cells. Expression
CC vectors comprising this cDNA can be transformed into host cells to
CC express GBS toxin receptor and its fragments. Detecting the receptor in
CC tissues is used to diagnose pathological vascularisation, e.g. for
CC detecting cancer metastases. GBS toxin receptors are useful for treating
CC neovascularisation (specifically cancer, reperfusion injury, scarring
CC during wound healing, keloids, chronic inflammation (rheumatoid
CC arthritis or psoriasis) or neural injury), and to raise specific
CC antibodies used for treating early onset disease. Inhibitors of this
CC receptor are useful for treating pathological or hypoxia-induced
CC endothelial cell proliferation and migration.
XX
SQ Sequence 2844 BP; 745 A; 623 C; 594 G; 882 T; 0 other;

Query Match 100.0%; Score 2844; DB 21; Length 2844;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2844; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccggggcgggggcggttcggcggtccgcgtgagctctctttccggcgagcaggttg 60
DB 1 cccggggcgggggcggttcggcggtccgcgtgagctctctttccggcgagcaggttg 60
QY 61 cgcgcgtagctccctgaagagcatcgaagtcgccggttcgagcttagcccgagcgagc 120
DB 61 cgcgcgtagctccctgaagagcatcgaagtcgccggttcgagcttagcccgagcgagc 120
QY 121 gcgagggaggtcgagccgcaacgcgtctctgcagcgcccgcgggcggaacgcgtc 180
DB 121 gcgagggaggtcgagccgcaacgcgtctctgcagcgcccgcgggcggaacgcgtc 180
QY 181 cagatgctctctgcgtttacaactagcatcttctgtcttttgggtttcttgctc 240
DB 181 cagatgctctgcgttcgtttacaactagcatcttctgtcttttgggtttcttgctc 240
QY 241 tctatcatcaggtgtaatctgagcgtltgactagtgagacatgltgattcaaaccaa 300
DB 241 tctatcatcaggtgtaatctgagcgtltgactagtgagacatgltgattcaaaccaa 300
QY 301 ctgcgaagaatagagaagctcctaagtggtgcagagcatctgtccataaagttc 360
DB 301 ctgcgaagaatagagaagctcctaagtggtgcagagcatctgtccataaagttc 360
QY 361 ttcccaaccgaagcggttaaaaagatcaggtggtgagatgagaaactcaaggtgattctcg 420
DB 361 ttcccaaccgaagcggttaaaaagatcaggtggtgagatgagaaactcaaggtgattctcg 420
QY 421 gatcttttctatgctacatcatcacacaatttccctggaagatagatgtgcccagcaa 480
DB 421 gatcttttctatgctacatcatcacacaatttccctggaagatagatgtgcccagcaa 480
QY 481 gtggtggggaagctgtgctagatctcggaatctttgctaaagcatcttcaacctgtca 540
DB 481 gtggtggggaagctgtgctagatctcggaatctttgctaaagcatcttcaacctgtca 540
QY 541 ctccccctgcgtcagattctcgaggtcgagcccttgtcactcaaggcactagaagggc 600
DB 541 ctccccctgcgtcagattctcgaggtcgagcccttgtcactcaaggcactagaagggc 600

QY 601 taagagaggtgtcacatcacagccatgcatgcatgtgcttcaatgggtcccccctc 660
DB 601 tagagagaggtgtcacatcacagccatgcatgcatgtgcttcaatgggtcccccctc 660
QY 661 ttgaagaagaagcctctgagatattcatatgcagaagcacaaacttggagcagttt 720
DB 661 ttgaagaagaagcctctgagatattcatatgcagaagcacaaacttggagcagttt 720
QY 721 ctctctctctctgagatatttgcactatgtatgtacttattgtcttattttc 780
DB 721 ctctctctctctgagatatttgcactatgtatgtacttattgtcttattttc 780
QY 781 ttgcatgttggaatcatctgttttaattcatagatctcttagttagtgcacaccag 840
DB 781 ttgcatgttggaatcatctgttttaattcatagatctcttagttagtgcacaccag 840
QY 841 aaactcaagaacaatcactccctatgaaaagagatattcttcatcattaaaaaac 900
DB 841 aaactcaagaacaatcactccctatgaaaagagatattcttcatcattaaaaaac 900
QY 901 agctcttcttacaagaatgcagtcggtgagatcctatgctgaatacactgtccattgg 960
DB 901 agctcttcttacaagaatgcagtcggtgagatcctatgctgaatacactgtccattgg 960
QY 961 ctattgtcgttgacatttcttcaaacctgagactttaaacttgttgaacttttg 1020
DB 961 ctattgtcgttgacatttcttcaaacctgagactttaaacttgttgaacttttg 1020
QY 1021 ctacttacaatgaagaagtccttaagtlcacaattccaagaagaatggttttatctgcag 1080
DB 1021 ctacttacaatgaagaagtccttaagtlcacaattccaagaagaatggttttatctgcag 1080
QY 1081 tcccttaattaggttgtgtgtatgtatgtacctgtgcgggttcaagctgcgaataattta 1140
DB 1081 tcccttaattaggttgtgtgtatgtatgtacctgtgcgggttcaagctgcgaataattta 1140
QY 1141 gggcaagatgaaatttctcaacacctgtgggttgcgaagagtttcttaacctatagagatga 1200
DB 1141 gggcaagatgaaatttctcaacacctgtgggttgcgaagagtttcttaacctatagagatga 1200
QY 1201 ttgagacctgcgatactccgtgttcgcgagagattataggtctgtattctcttgctg 1260
DB 1201 ttgagacctgcgatactccgtgttcgcgagagattataggtctgtattctcttgctg 1260
QY 1261 ttgattccctaacaacatatacaaacacctggagagcttcttccttctgattagataca 1320
DB 1261 ttgattccctaacaacatatacaaacacctggagagcttcttccttctgattagataca 1320
QY 1321 accatctggaatctgtcctctgtatgtgttattctcctggtgcatacaaaatcccttg 1380
DB 1321 accatctggaatctgtcctctgtatgtgttattctcctggtgcatacaaaatcccttg 1380
QY 1381 ccaactatccctggaatgattgggccatcaatttgcagaagttcttaacccctgagaacacta 1440
DB 1381 ccaactatccctggaatgattgggccatcaatttgcagaagttcttaacccctgagaacacta 1440
QY 1441 ttgagaaatggcaaacctgttttctgcatcgctgtccttcaatgattttggtgcaatt 1500
DB 1441 ttgagaaatggcaaacctgttttctgcatcgctgtccttcaatgattttggtgcaatt 1500
QY 1501 tcttcaacactatcgcgcaagtgaaagtccaacacggggccatcagtgatccacaagagac 1560
DB 1501 tcttcaacactatcgcgcaagtgaaagtccaacacggggccatcagtgatccacaagagac 1560
QY 1561 acagaacttgaaggaaccaataataatccctgtctctatthaatglatcttgtttatcat 1620
DB 1561 acagaacttgaaggaaccaataataatccctgtctctatthaatglatcttgtttatcat 1620
QY 1621 gtaactaaagtgcctttgtatatttaatgttgaagcaatctatataaagaataaat 1680
DB 1621 gtaactaaagtgcctttgtatatttaatgttgaagcaatctatataaagaataaat 1680

QY 1681 gtaataaagaaattgtctagatcttgtaaggcttgtaacatgaataatgtaactagtgtcc 1740
|||||
Db 1681 gtaataaagaaattgtctagatcttgtaaggcttgtaacatgaataatgtaactagtgtcc 1740
QY 1741 atataagcaaaatagactattttaaattattataacccgttgcgtgaacttaacatcc 1800
|||||
Db 1741 atataagcaaaatagactattttaaattattataacccgttgcgtgaacttaacatcc 1800
QY 1801 aggtacataatctgtcgaactgaactgaaccccaacaaaggaaggagttattattat 1860
|||||
Db 1801 aggtacataatctgtcgaactgaactgaaccccaacaaaggaaggagttattattat 1860
QY 1861 aagaccataacctaaagaagaatgaactgaataatagacccttcataccttgccttaataag 1920
|||||
Db 1861 aagaccataacctaaagaagaatgaactgaataatagacccttcataccttgccttaataag 1920
QY 1921 tggataataatctcgaactgtctgttaaacatctgttttgcataccttcctcaaaaatt 1980
|||||
Db 1921 tggataataatctcgaactgtctgttaaacatctgttttgcataccttcctcaaaaatt 1980
QY 1981 attgtacataacatccctgacatgaactgaactgaactgaactgaactgaactgaactga 2040
|||||
Db 1981 attgtacataacatccctgacatgaactgaactgaactgaactgaactgaactgaactga 2040
QY 2041 cagccactgtatcatcactgaactgaactgaactgaactgaactgaactgaactgaactga 2100
|||||
Db 2041 cagccactgtatcatcactgaactgaactgaactgaactgaactgaactgaactgaactga 2100
QY 2101 acatgtccctctctgtctcaaggacagtcccaagcaactgaactgaactgaactgaactga 2160
|||||
Db 2101 acatgtccctctctgtctcaaggacagtcccaagcaactgaactgaactgaactgaactga 2160
QY 2161 ggtcaagcgcgaaggcttgtagacgtatcttccctggggcgttttaattgtgtgagaag 2220
|||||
Db 2161 ggtcaagcgcgaaggcttgtagacgtatcttccctggggcgttttaattgtgtgagaag 2220
QY 2221 cccctgagccaacagagagcagcgatcccaagatcagtttccatcacccctccctc 2280
|||||
Db 2221 cccctgagccaacagagagcagcgatcccaagatcagtttccatcacccctccctc 2280
QY 2281 cctcccaagcaactggaagtattgtcgtgacatgaactgaactgaactgaactgaactga 2340
|||||
Db 2281 cctcccaagcaactggaagtattgtcgtgacatgaactgaactgaactgaactgaactga 2340
QY 2341 aattgacacataatacatcatccttgatgatccttcacatccttcacatccttcacatc 2400
|||||
Db 2341 aattgacacataatacatcatccttgatgatccttcacatccttcacatccttcacatc 2400
QY 2401 tctcaagtcctgtgtcttagagagctacacacagaaataataataaagaagaacaa 2460
|||||
Db 2401 tctcaagtcctgtgtcttagagagctacacacagaaataataataaagaagaacaa 2460
QY 2461 acatttttctgtctgacacctaagtgtctgtgtcagttcaagttaaagttaacttcta 2520
|||||
Db 2461 acatttttctgtctgacacctaagtgtgtgtgtcagttcaagttaaagttaacttcta 2520
QY 2521 ctccacataacccgcaacccggtgtgttaacatcctttagtgtgtgttcttaattttat 2580
|||||
Db 2521 ctccacataacccgcaacccggtgtgttaacatcctttagtgtgtgttcttaattttat 2580
QY 2581 tttccagtttctcctgagacatctccagttgttttgagcagtttgagggcattat 2640
|||||
Db 2581 tttccagtttctcctgagacatctccagttgttttgagcagtttgagggcattat 2640
QY 2641 gttgattgaactgtatgaatgtttcaactgtgtatgttcaacactacgttaatttta 2700
|||||
Db 2641 gttgattgaactgtatgaatgtttcaactgtgtatgttcaacactacgttaatttta 2700
QY 2701 actaaagcatttaagtataatgaatgtatgatccttcgttaaatatttttaactcg 2760
|||||
Db 2701 actaaagcatttaagtataatgaatgtatgatccttcgttaaatatttttaactcg 2760
QY 2761 taaatagcttaagtgtcgtatgtatcttcttcaacaatcaataaatcttttgg 2820

Db 2761 taaatagcttaagtgtcgtatgtatcttcttcaacaatcaataaatcttttgg 2820
QY 2821 aatgataaaaaaaaaaaaaaaaaa 2844
Db 2821 aatgataaaaaaaaaaaaaaaaaa 2844
RESULT 2
AAD10326
ID AAD10326 standard; DNA; 2844 BP.
XX
AC AAD10326;
XX
DT 16-OCT-2001 (first entry)
XX
DE Sheep group B beta-haemolytic Streptococci toxin receptor (SP55) DNA.
XX
KW Sheep; group B beta-haemolytic Streptococci toxin receptor; GBS; SP55;
KW cytostatic; vulnereary; antiatherosclerotic; osteopathic; vasotrophic;
KW prevention; attenuation; pathoangiogenic condition; cancer; scar;
KW wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;
KW keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;
KW vaccine; ds.
XX
OS Ovis sp.
XX
FH Key Location/Qualifiers
FT CDS 84..1571
FT /tag="a
FT /product="Sheep GBS toxin receptor protein (SP55)"
XX
PN MO200156598-A2.
XX
PD 09-AUG-2001.
XX
PF 02-FEB-2001; 2001MO-US03662.
XX
PR 02-FEB-2000; 2000US-0179870.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
PI
XX
PI
XX
DR MPI: 2001-488844/53.
DR P-PSDB: AAE06519.
XX
PT Preventing or attenuating pathoangiogenic conditions e.g. cancer,
PT chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by
PT administering group B beta-hemolytic Streptococci toxin receptor or its
PT fragment -
XX
XX
PS Disclosure: Page 47-50; 52pp; English.
XX
XX
CC The present sequence is a DNA encoding sheep group B beta-haemolytic
CC Streptococci (GBS) toxin receptor protein, SP55. The present invention
CC relates to a method for preventing or attenuating a patho-angiogenic
CC condition in a mammal which comprises administering to the mammal one
CC or more GBS toxin receptors or their immunogenic fragments to induce
CC or maintain an immune response to one of GBS toxin receptors. The
CC method is useful for preventing or ameliorating pathoangiogenic
CC conditions such as cancer, scarring during wound healing, gliosis
CC during repair of nerve injury, chronic wounds, keloids, reperfusion
CC injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and
CC psoriasis in mammals. The proteins of the invention are also used
CC as vaccines.
XX
SQ Sequence 2844 BP; 745 A; 623 C; 594 G; 882 T; 0 other;

Query Match 100.0%; Score 2844; DB 22; Length 2844;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2844; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 cccggggcgggggcttcgagcgtcccgctggagctctcttccggggagcaggtttg 60
Db 1 cccggggcgggggcttcgagcgtcccgctggagctctcttccggggagcaggtttg 60
QY 61 cggcgtagctccctgaaggcatcatgaagtcgccggtttcgagcttagcccgagcgag 120
Db 61 cggcgtagctccctgaaggcatcatgaagtcgccggtttcgagcttagcccgagcgag 120
QY 121 gcgaggagggtcgagcgcgcacgcgctcctcgagcgcgcccgcgcgagaaacccgctc 180
Db 121 gcgaggagggtcgagcgcgcacgcgctcctcgagcgcgcccgcgcgagaaacccgctc 180
QY 181 cagatagctgcctgcctgtctacacaaactagcattttgctcttttggtttcttcgtc 240
Db 181 cagatagctgcctgcctgtctacacaaactagcattttgctcttttggtttcttcgtc 240
QY 241 tctatctaatcgggtgaatctgagcgtttgacactagtgagacatggtgatccaacaa 300
Db 241 tctatctaatcgggtgaatctgagcgtttgacactagtgagacatggtgatccaacaa 300
QY 301 ctgccaagaataatagaaacgtccacagagtgtagagagcatttcgtccccaataagttc 360
Db 301 ctgccaagaataatagaaacgtccacagagtgtagagagcatttcgtccccaataagttc 360
QY 361 ttccaacccaacacggttaaaaagtacccgttggtga tgcagaaactcaagatgga tttctcg 420
Db 361 ttccaacccaacacggttaaaaagtacccgttggtga tgcagaaactcaagatgga tttctcg 420
QY 421 gatctttttctatgctacatcatcaacaacttcctgagagatagtggtcgagcaaa 480
Db 421 gatctttttctatgctacatcatcaacaacttcctgagagatagtggtcgagcaaa 480
QY 481 gtgggggggaagcgtgtgctaagatctggagatctgtctacagcattcttcaacccgtttca 540
Db 481 gtgggggggaagcgtgtgctaagatctggagatctgtctacagcattcttcaacccgtttca 540
QY 541 ctccctcctgcgtgcagatcttcggaatctcgagccctgtgtgcactcagagcgacataagggc 600
Db 541 ctccctcctgcgtgcagatcttcggaatctcgagccctgtgtgcactcagagcgacataagggc 600
QY 601 taggaagagggtgtacacatataccacgcatgcatggtgtctctcataggtcccccctc 660
Db 601 taggaagagggtgtacacatataccacgcatgcatggtgtctctcataggtcccccctc 660
QY 661 ttgaaagaagaagcctctcgatatttcacatagcagagcacaactcggggcagtagttt 720
Db 661 ttgaaagaagaagcctctcgatatttcacatagcagagcacaactcggggcagtagttt 720
QY 721 cctctcctcttctggaatattgtgctactataatgaatltggaactatgctctctattct 780
Db 721 cctctcctcttctggaatattgtgctactataatgaatltggaactatgctctctattct 780
QY 781 ttggcattgtttggatacatctggtttattttatgatactgcttagtgatagatcacacag 840
Db 781 ttggcattgtttggatacatctggtttattttatgatactgcttagtgatagatcacacag 840
QY 841 aaactcaagaacatacctccctatgaaaaaggatatactcttcatcatataaatac 900
Db 841 aaactcaagaacatacctccctatgaaaaaggatatactcttcatcatataaatac 900
QY 901 agctctcttcacagaagtcagtgccgtggaatacctatgctgtaaaactcgaactctggg 960
Db 901 agctctcttcacagaagtcagtgccgtggaatacctatgctgtaaaactcgaactctggg 960
QY 961 ctattgtcgttgacaattttcttaacaactgagacttttatacttvtgtgaacctatgct 1020
Db 961 ctattgtcgttgacaattttcttaacaactgagacttttatacttvtgtgaacctatgct 1020
QY 1021 ctacttacaagaagaagtccttaaggttcaatatcaagaagaatgggttttattctgag 1080
Db 1021 ctacttacaagaagaagtccttaaggttcaatatcaagaagaatgggttttattctgag 1080

QY 1081 tcccttattaggtgtgtgataatgataatcctcgtcgggtcaagcgtcgtgacaatttaa 1140
Db 1081 tcccttattaggtgtgtgataatgataatcctcgtcgggtcaagcgtcgtgacaatttaa 1140
QY 1141 gggcaagaatggaattttcaacactctgtggttgcgaagatltttagcctataagagata 1200
Db 1141 gggcaagaatggaattttcaacactctgtggttgcgaagatltttagcctataagagata 1200
QY 1201 ttggaaccttggaatttctgtgttcggcagaagatttaagcgtggaattatctctgtgtc 1260
Db 1201 ttggaaccttggaatttctgtgtgttcggcagaagatttaagcgtggaattatctctgtgtc 1260
QY 1261 ttgcatcctcaataatcacaacacccctgggagagcttttctctctcgtgatttagcata 1320
Db 1261 ttgcatcctcaataatcacaacacccctgggagagcttttctctctcgtgatttagcata 1320
QY 1321 accatctggaacattgctccttcgtatgctgtaattcctcgtggcatcaacaatacctttg 1380
Db 1321 accatctggaacattgctccttcgtatgctgtaattcctcgtggcatcaacaatacctttg 1380
QY 1381 ccaacttctcgtgaatgatttggggcccatcttgccagaagcttcaacccctgagaacacta 1440
Db 1381 ccaacttctcgtgaatgatttggggcccatcttgccagaagcttcaacccctgagaacacta 1440
QY 1441 ttggaagaatggaacactgttttctgtcatcgtcgtgctatcaaatglatttgtgtccatt 1500
Db 1441 ttggaagaatggaacactgttttctgtcatcgtcgtgctatcaaatglatttgtgtccatt 1500
QY 1501 tcttcaacatacttcgccaagggtgaagtgcataaaactgtggccaatgataccaagagac 1560
Db 1501 tcttcaacatacttcgccaagggtgaagtgcataaaactgtggccaatgataccaagagac 1560
QY 1561 acagaacactgaaggaaccaataataatcctgtctcatttaagtattcctgtttattcat 1620
Db 1561 acagaacactgaaggaaccaataataatcctgtctcatttaagtattcctgtttattcat 1620
QY 1621 gtaaccttaaaagtgcctttgataattttaatgtgtgaagcaatctatacaagaataaatt 1680
Db 1621 gtaaccttaaaagtgcctttgataattttaatgtgtgaagcaatctatacaagaataaatt 1680
QY 1681 gtaactaagaaaaatgtgttgaatttgaagcgtgtgaatacatgaagaatgtaactagtgc 1740
Db 1681 gtaactaagaaaaatgtgttgaatttgaagcgtgtgaatacatgaagaatgtaactagtgc 1740
QY 1741 atataagcaaaatagctatttttaattatatttaaacccgtttgtggaacttaacaattc 1800
Db 1741 atataagcaaaatagctatttttaattatatttaaacccgtttgtggaacttaacaattc 1800
QY 1801 aggttcacatactgctgcaagtcaggaagcaaccacaataagggagttctattattat 1860
Db 1801 aggttcacatactgctgcaagtcaggaagcaaccacaataagggagttctattattat 1860
QY 1861 aagaccataactaagaagatgagctggaatagacccttataccttggcttaattaag 1920
Db 1861 aagaccataactaagaagatgagctggaatagacccttataccttggcttaattaag 1920
QY 1921 ttgataataattcgaagctgtgttaaacatctgtttttgaacacttccccaataaatt 1980
Db 1921 ttgataataattcgaagctgtgttaaacatctgtttttgaacacttccccaataaatt 1980
QY 1981 attgtatcagaacatccctgacatgtagtltccaacacttagccttccaagagagctgg 2040
Db 1981 attgtatcagaacatccctgacatgtagtltccaacacttagccttccaagagagctgg 2040
QY 2041 cagccactgtatacatcgaagccttgcaacttcaactgaggaagacatgcccagagcgtgcc 2100
Db 2041 cagccactgtatacatcgaagccttgcaacttcaactgaggaagacatgcccagagcgtgcc 2100
QY 2101 acatgtccctctcctgcttcaggaagtcgccaagacacttaaggcaagcaccagagcag 2160
Db 2101 acatgtccctctcctgcttcaggaagtcgccaagacacttaaggcaagcaccagagcag 2160
QY 2161 gtcagcgccaagcctttagacgtaattctcccttggggcgtttaatgtgtgatatgaag 2220
```


| | | | |
|----------|--|--|------|
| Db | 2161 | ggtcaagcgcaaggcctttgtagcggtatctctcccgggcgctgtaatggtgtgtagag | 2220 |
| OY | 2221 | ccctgaagccaacagggacgcgcgatccacagtcagtgttccatgcaacctctccctc | 2280 |
| Db | 2221 | ccccgagccaacagggacgcgcgatccacagtcagtgttccatgcaacctctccctc | 2280 |
| OY | 2281 | ccttcacagacacactgagatgttgcctgtagttaaaccgcgaanaagaaagtgtgtagct | 2340 |
| Db | 2281 | ccttcacagacacactgagatgttgcctgtagttaaaccgcgcgaanaagaaagtgtgtagct | 2340 |
| OY | 2341 | aattagccacataatacatctcttgatgatctcccttcacatgagatcagagatataa | 2400 |
| Db | 2341 | aattagccacataataatacatctcttgatgatctcccttcacatgagatcagagatataa | 2400 |
| OY | 2401 | tcttcaagtcctggtgttctagtagctacacacagaataataataataaagaacacaa | 2460 |
| Db | 2401 | tcttcaagtcctggtgttctagtagctacacacagaataataataataaagaacacaa | 2460 |
| OY | 2461 | acattttctctgtctgcgacccaagtgtctggtgttcagttcaagttaaagttagctcta | 2520 |
| Db | 2461 | acattttctctgtctgcgacccaagtgtctggtgttcagttcaagttaaagttagctcta | 2520 |
| OY | 2521 | cttcaacataccgacacccggtgtgttaatacatttaagtgtgtttccttaaccttat | 2580 |
| Db | 2521 | cttcaacataccgacacccggtgtgttaatacatttaagtgtgtttccttaaccttat | 2580 |
| OY | 2581 | tttccaggttttccctgacacatctccacagtggtgttttgagcagctgtcttgaggacattat | 2640 |
| Db | 2581 | tttccaggttttccctgagacacatctccacagtggtgttttgagcagctgtcttgaggacattat | 2640 |
| OY | 2641 | gtgatttagaactgtaataagtgttcaactgtgtatgttcaacactacactgtaataattta | 2700 |
| Db | 2641 | gtgatttagaactgtaataagtgttcaactgtgtatgttcaacactacactgtaataattta | 2700 |
| OY | 2701 | actaaagctatttaatgtatataatgtagtatacatctcgttaataatttttaactctg | 2760 |
| Db | 2701 | actaaagctatttaatgtatataatgtagtatacatctcgttaataatttttaactctg | 2760 |
| OY | 2761 | taaatagctttaagttgctatgtgtgtagatattcttttaacaatccaataaataatcctttcg | 2820 |
| Db | 2761 | taaatagctttaagttgctatgtgtgtagatattcttttaacaatccaataaataatcctttcg | 2820 |
| OY | 2821 | aatgataaaaaaaaaaaaaaaaaa 2844 | |
| Db | 2821 | aatgataaaaaaaaaaaaaaaaaa 2844 | |
| RESULT 3 | | | |
| AAF55900 | | | |
| ID | AAF55900 standard; DNA; 2512 BP. | | |
| XX | AAF55900; | | |
| AC | 18-APR-2001 (first entry) | | |
| XX | Human AST coding sequence. | | |
| DT | Human; AST; noctropic; immunotropic; gene therapy.. Salla disease; | | |
| XX | anion and sugar transporter; anion-cation symporter; | | |
| KW | Salla acid transporter; ss. | | |
| KM | Homo sapiens. | | |
| XX | EPI069184-A1. | | |
| PN | 17-JAN-2001. | | |
| XX | 16-JUL-1999; 99EP-0202341. | | |
| PF | 16-JUL-1999; 99EP-0202341. | | |
| XX | (ALKU) AKZO NOBEL NV. | | |

| | |
|----|--|
| XX | WPI: 2001-193090/20. |
| DR | P-PSDB: AAB66967. |
| XX | |
| PT | New human transporter gene implicated in Salla disease and lysosomal |
| PT | sialic acid transport, useful in assays for identifying new drugs, or |
| PT | diagnosing sialic acid transport defects related to mutations in the |
| PT | transporter gene - |
| XX | |
| PS | Claim 2; Page 12-13; 20pp; English. |
| XX | |
| CC | The present sequence is the coding sequence for human Anion and Sugar |
| CC | Transporter (AST) protein. AST has significant homology with several |
| CC | members of the anion-cation symporter (ACS) family of transporters. AST |
| CC | is implicated in Salla disease, and is useful in screening assays for |
| CC | identifying new drugs. Compounds identified via AST screening is useful |
| CC | for preparing a pharmaceutical suitable as an activator or inhibitor of a |
| CC | sialic acid transporter protein. The pharmaceutical may be used in sialic |
| CC | acid associated diseases and CNS/immune related disorders. |
| XX | |
| SQ | Sequence 2512 BP; 612 A; 585 C; 566 G; 749 T; 0 other; |
| | |
| | Query Match 55.3%; Score 1572.4; DB 22; Length 2512; |
| | Best Local Similarity 82.1%; Pred. No. 0; |
| | Matches 1917; Conservative 0; Mismatches 386; Indels 33; Gaps 8 |
| OY | 5 ggggagggggggttcgcgcgtcccgctcgtgagctcttcttcgcggagcagtgttgccc 64 |
| Db | 194 ggcgcggggggtcgtggctgtctggcgcggcgctccctcttcgccaggtgvcgaatacac 253 |
| OY | 65 gtatgcctccgaagaagcatcaatgaatccccggatttcgacttagccccgagcagcgca 124 |
| Db | 254 ct-gctacagtaagcgtcataggtcttcggttcgaacctgcggcccggaacgatvgcca 311 |
| OY | 125 ggagggtctcgacgcgcacacacgcgtctctgagcgcgcgcggcgggcggaaccgcgtccat 184 |
| Db | 312 ggaagacacgcgcgcgcacgcgtctcttaccagggcgccccaagcgacgaagcgcgtccagt 371 |
| OY | 185 atgcgcctcgtctgttacaaccttgaatcttcttgccttttgtgttcttcgtcttca 244 |
| Db | 372 gtgcgcctcgtcgtcttaacaacttgcaatttgccttcttggtttcttcaattgltga 431 |
| OY | 245 ttcatlaccggtgaactctgagcgtgtgcactagtgacatggttggtttcaaacacaactgc 304 |
| Db | 432 tgcattacgctggaatctgagtggtgtgcttagtgatatgtagattcaatacaactt 491 |
| OY | 305 caaagataatagaacgctcctaagatgtfvcagagcaatctctcccataaaagtcttca 364 |
| Db | 492 agaagataatagaacttccaaagcgtgtccagagacattctgtccatataaagttalca 551 |
| OY | 365 caaccaaaaacggtaaanaagtaacggttgagatccagaanaactaagatgatcttcgcagc 424 |
| Db | 552 taatlcaaaacggtaagaagtlaccaatggagatgcagaactcaagatgattcttcggttc 611 |
| OY | 425 ttcttctcatggtcatcatcatcacacaacttccttgtaggataatgttgcacagagaatgg 484 |
| Db | 612 ctttttttatggtcatcatcatcaacaacagatttccgtggagatatgtttgcagcaaaaatagg 671 |
| OY | 485 ggggaagcgtgtgtcagatctcggaatctttgtctacagctatcttccacccgttcaatcc 544 |
| Db | 672 ggggaaaatgctgcctagagatcttggaatccttgcgaactgcttccataccctgtcaatcc 731 |
| OY | 545 cctgcgtcagatcttcgagatgctggagcctgtttgcactcagggacaggaagggctagg 604 |
| Db | 732 catgtcctcagatttagaagatgtgacacactcatgtlaactcagaagacactagaagaagctagg 791 |
| OY | 605 aagaaggttcaacatactcagcatcatgacatgacatgtgtcttcatgaggtccccccttta 664 |
| Db | 792 agaggggtttacattcttcagccatgcatgccaatgtggtctctcttgggtccccctcttga 851 |
| OY | 665 aaagaagcaagctcttcagatatttcalatgacaggaacacaacttggagacagtattctct 724 |

Db 852 aagaagcaaacctcttagcatcttcataatgacagagacacagcttgggacgaattctct 911
 QY 725 tccctctctcggagtaattgctactatacgaaatgagactatgctctcattctctgg 784
 Db 912 tccctctctcggagtaattgctactatacgaaatgagactatgctctcattctctgg 971
 QY 785 catgttggagatacatctggtttattatgacatcgtcttaagttagtatacaccagaac 844
 Db 972 taactatgagaatttttggcttcttcttggatctggttagttagtgaacacacacaaa 1031
 QY 845 tcaacagacatacaccctcgtataaaaggagttatctcttcacgttaaaanaatcagct 904
 Db 1032 accaaagagaattcccatatgaaaaggaaatacattcttcacatlaagaaacgcgt 1091
 QY 905 cctctacagaagtcagtcgcgtgatacctatgctgaatgaactgcacacttgggctat 964
 Db 1092 tctctacagaagtcagtcgcgtgatacctatgctgaatgaactgcacacttgggctat 1151
 QY 965 tgcgttgcacatttcttacaactggacttttatacttggttgaacctattgacctac 1024
 Db 1152 cgtagtgcacacttcttacaactggacttttatacttgaacattattgacctac 1211
 QY 1025 ttacatgaagaagtcagtcggtacatatacagaagaatgggtttttatctgcagctcc 1084
 Db 1212 ttatagaagaagacaccgaagtlcaatgttcaagaagaatgggtttttatctgcagctcc 1271
 QY 1085 ttattagtgctgtgtatgatacctcgtcgtgcagtcgacagtcgacacattaaagagc 1144
 Db 1272 ttattagtgctgtgtatgatacctcgtcgtgcagtcgacagtcgacacattaaagagc 1331
 QY 1145 aagaatgaatttcaactcgtggttcgaagaatgttttagcctataagagatgttg 1204
 Db 1332 aaaaatgaatttcaactcgtggttcgaagaatgttttagcctataagagatgttg 1391
 QY 1205 acctgcgatacttcctggttcgcgagagatattagagctgagatatttcctcgtctggc 1264
 Db 1392 acctgcgatacttcctggttcgcgagagatattagagctgagatatttcctcgtctggc 1451
 QY 1265 attcctaaccatacacaacaccttggaggcttctcctctcgtatlaagacataacca 1324
 Db 1452 ttctctaaccatacacaacaccttggaggcttctcctcgtatlaagacataacca 1511
 QY 1325 tctgcgacattgctccttgcgtatgctgtatctcctcgtggacatacaaatccttggcac 1384
 Db 1512 tctgcgacattgctccttgcgtatgctgtatctcctcgtggacatacaaatccttggcac 1571
 QY 1385 tattcctggaatgattgggcccacatctgcagaaagttcaacccctggagaacatttgg 1444
 Db 1572 tattcctggaatgattgggcccacatctgcagaaagttcaacccctggagaacatttgg 1631
 QY 1445 agaatggcaaacgttcttgcacatcgtcgtatcaatgataatttgggtgacatttctt 1504
 Db 1632 agaatggcaaacgttcttgcacatcgtcgtatcaatgataatttgggtgacatttctt 1691
 QY 1505 cacactattgcgcaaaagtgaagtgaacacatgggcccacatgagtcacacaaagagacag 1564
 Db 1692 tacactattgcgcaaaagtgaagtgaacacatgggcccacatgagtcacacaaagagacag 1751
 QY 1565 aaactgaagaaacaaataataatcctgtcctataatgatacttggttatcatagttaa 1624
 Db 1752 aaactgaagaaacaaataataatcctgtcctataatgatacttggttatcatagttaa 1811
 QY 1625 cctaaagtcgaccttgataatttaagtgtgaagcaatcatatacaagaataaattgttac 1684
 Db 1812 cctcaaaagtcgacctt-----ctgtatgtgtgaagcattcatagt-----cttttct 1857
 QY 1685 tagaaaaattgtgtagattgtgaagcgtgtgaatcaatgaatgtcaatgtgtccatat 1744
 Db 1858 aatgtgacctgttagatttttaagcctataatcatcatgaatatacatcagtgccagaa 1917
 QY 1745 aagcaaaatagctatttttaataatataaaccogtctgctggaactcaaatcagag 1804
 Db 1918 taataaaatgaactgtcttaataatgaataatgaatgaatgaatgaatgaatgaatgaat 1977

QY 1805 tcaacatactgctgcaagtcagacacaccacataagaggagttctcatla-ttataag 1863
 Db 1978 tcaacataccgctcgtcagtcgagcaacatagaaagtagaaagtcgttgcatttttag 2037
 QY 1864 accataactaagaagatgagctgaataagaccccttataaccttgccttaataaagttg 1923
 Db 2038 gccataactaagaagatgagctgaataagaccccttgccttaataaactag 2097
 QY 1924 atataatctcaggtctgtttaaacaatcgttcttgaacacttcccaaaaattatt 1983
 Db 2098 atgataatctcaggtctgtttaaacaatcgttcttgaacacttcccaaaaattatt 2152
 QY 1984 tgcatacagaacacccctgacatgtagtctcaaaactttagcctctcagagagctggcag 2043
 Db 2153 -atgtgcagctcctctgacacttagaccccaactttagcctctcagagagctggcag 2211
 QY 2044 ccactgtataatcagcctgcaacttcaactgaggaagacatgccagagctggcaca 2103
 Db 2212 ccactgtataatcagcctgcaactgagctgagggagtggtgccagagctggcaca 2271
 QY 2104 tgcacctcctgctgctcagggagacatgcccagcactt-----agcagacatcaaga 2156
 Db 2272 cactccctcctgctcagggagcagatgcccagcacttatacagagcagacatcaaga 2331
 QY 2157 ccagggtcagcgcgaagccttggacggtatcttccctc-9gggctgttaatgtgtg 2214
 Db 2332 ccagggtcagcgcgtcagcctcttgcgtgctgcttccctcctcagggagcctatgtag 2391
 QY 2215 atgaagccctgagcacaagggagcagcagatcacaatcagatgttctcatgacccctc- 2273
 Db 2392 atgaagccctgagcacaagggagcagcagatcacaatcagatgttctcatgacccctc 2451
 QY 2274 tcccttcccttcccaacacactgagatgcttgcgtgacatgacccctgcaaaagaa 2329
 Db 2452 aacttcccttcccaacacagagaatatgtgctgagcagcaccctgcaaaagaa 2507

RESULT 4
 ID AA250879 standard; cDNA: 2930 BP.
 XX
 AC AA250879;
 XX
 DT 31-MAY-2000 (first entry)
 XX
 DE Full length human GBS toxin receptor (HP59) cDNA.
 XX
 KW Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59;
 KW pathological vascularisation; cancer metastases; angiodermatitis;
 KW neovascularisation; reperfusion injury; scarring; keloid;
 KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
 KW endothelial cell proliferation; antibacterial; anticancer;
 KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 263..1873
 FT /tag= a
 FT /product= "Human GBS toxin receptor protein"
 PN MO200005375-A1.
 XX
 PD 03-FEB-2000.
 XX
 PF 22-JUL-1999; 99WO-DS16676.
 XX
 PR 22-JUL-1998; 98US-0093843.
 XX
 PA (UYVA-) UNTV VANDERBILT.
 XX
 PI Helleqvist CG, Fu C;

847 cattgtgcagattlaagsgtltggnaccatcttgtactcagaagcactagaagactagg 906
QY 605 agagggtgcacatataccagcatgcatgcatgtgtcttcattggtccccccttga 664
DB 907 agagggtgttacctatctccagcatgcatgcatgtgtgtcttcttggcccccttga 966
QY 665 aagaagaagcctctgagttattcatatgcaagagacaaacttggagacagttctct 724
DB 967 aagaagaacactcttagattctgtatgcagagacagccttggagacagtaattctct 1026
QY 725 tccctcttgcaggaattgtctactatataatgtgacttatgtcttattcttcttg 784
DB 1027 tccctcttgcaggaattgtctactatataatgtgacttatgtcttattcttcttg 1086
QY 785 cattgtgcagcatctgtgtttatttatgtatcgtctgagttagttagtacaacgaac 844
DB 1087 tactattggaattatttgttcttcttcttcttcttcttcttcttcttcttcttct 1146
QY 845 tcaacaagaacatcaactcccatatgaaagagatatacttctcatatnaaaatcagct 904
DB 1147 acacaagaagaattcccatatgaaaggaatacattcttcatatnaaaatcagct 1206
QY 905 ccttcacagaagtcagtgccgtgtgatacctatgtcgaatacactgcacatttggctat 964
DB 1207 tcttcacagaagtcagtgccgtgtgatacctatgtcgaatacactgcacatttggctat 1266
QY 965 tgtctgtgcacatttcttctacactgagcttttataattgttgcacatttgcctac 1024
DB 1267 cgtagtgtgcacatttcttctacactgagcttttataattgttgcacatttgcctac 1326
QY 1025 ttacatgaagaagtccttaagttcaatatcaagaagaatgggttttactgcagctcc 1084
DB 1327 ttaatagaagaagtccttaagttcaatatcaagaagaatgggttttactgcagctcc 1386
QY 1085 ttatttaggtgtgtgtatgatatatcctgtccgggtcgaagtcgtcgaataattgaagc 1144
DB 1387 ttatttaggtgtgtgtatgatatatcctgtccgggtcgaagtcgtcgaataattgaagc 1446
QY 1145 aagatggaattttcaactcgtgtgtcgaagatttttaagcctatagaagatttg 1204
DB 1447 aagatggaattttcaactcgtgtgtcgaagatttttaagcctatagaagatttg 1506
QY 1205 acctgcgatattcctgtgtgtccgcaagattataagcgtgtgattatctcctgtgtg 1264
DB 1507 acctgcgatattcctgtgtgtccgcaagattataagcgtgtgattatctcctgtgtg 1566
QY 1265 attcctaaccatataaaacacctgtggaggtcttgccttcttgattagaacaaca 1324
DB 1567 ttctctaaccatataaaacacctgtggaggtcttgccttcttgattagaacaaca 1626
QY 1325 tctggaacattgtcctctgtatgctggtatctcctgtggaacacaaataacttgcac 1384
DB 1627 tctggaacattgtcctctgtatgctggtatctcctgtggaacacaaataacttgcac 1686
QY 1385 tattcctggaatgtgggcccacatctgtccagaagcttlaacccctggaacactatgg 1444
DB 1687 tattcctggaatgtgggcccacatctgtccagaagcttlaacccctggaacactatgg 1746
QY 1445 agaattggaacactgttctctgtatgctggtatctcctgtggaacacaaataacttgcac 1504
DB 1747 agaattggaacactgttctctgtatgctggtatctcctgtggaacacaaataacttgcac 1806
QY 1505 cacactatctgcgaagaagtgaagtgaagaacactggaacacacacacacacacacac 1564
DB 1807 tacactatctgcgaagaagtgaagtgaagaacactggaacacacacacacacacacacac 1866
QY 1565 aacatggaagaac 1624
DB 1867 aacatggaagaac 1926
QY 1625 ccttaagaatgcttctgtatatttaattgtgtgaagcactatataacaagaataaattgtgac 1684
DB 1927 ccttaagaatgcttctgtatatttaattgtgtgaagcactatataacaagaataaattgtgac 1972

QY 1685 tagaaatgtgttagatttgaagctgtgtaatcatgaataatgtcaagtgtccat 1744
DB 1973 aattgtactgttattagatttttaagggcctataatcaatgaataatcagttgccaga 2032
QY 1745 aagcaaatatgataatttaattatataaaccgttctgtcgaacttacaattcagag 1804
DB 2033 taataaagaactgtgttaattatgaataatgtgaagcagttcgttattttagt 2092
QY 1805 ttacatatctgtcgaagtcaagtcgaacccaataatagggagtctattta- ttataag 1863
DB 2093 tcaatacctgtcgttagtcgggaacacatgaatgaagcagttcgttattttagt 2152
QY 1864 accataccataagagatgtgcgaataagaacctcttaacctttagtcaatgaagtg 1923
DB 2153 gccataactaaaggaatgtgcgaacacgaacctcctcgtataaccttgttaattaaactag 2212
QY 1924 ataataattctcaggtctgtttaaacatctgttltgtacacacttccataaaattat 1983
DB 2213 atgataattctcaggttagtataaacactgtgtgttcaacttccataaa- ---- 2267
QY 1984 tgtcatcagaacatccctgacatgtagtgctcaaaactttagccttccacgagctggcag 2043
DB 2268 -atgtcagctctctctgacacttagaccataactttagcatctctgtgagctggcat 2326
QY 2044 ccactgtatcattcagcctggaacttcaatgagggaagcagccagcagctgccaca 2103
DB 2327 ccactgtatcattcagcctggaacttcaatgagggaagcagcagctgccaca 2386
QY 2104 tglccctctctgtctcagggacagtgcccgacact- ---- aggaagcatccaaga 2156
DB 2387 cactccctcctgtctcagggacagtgcccgagcttcatcagagagcatccaagc 2446
QY 2157 cgaagtgacggccaagcttggacggtattctccct- -gggggtgttaattgttg 2214
DB 2447 cgaagtcagtgctgactcttgcgtgtgccttctcctgtgaggggtcatcaatgttag 2506
QY 2215 atgaagccctggaacacagggacgagcgcgatccacagatcaggtttccatgcacctc- 2273
DB 2507 ataaagccctggtatggaagacagtgagatccactgtagtgcgttatcatcactca 2566
QY 2274 tcccttccctccacgaactgtgattgtcctgtgcagtataactgcgaagaaga 2329
DB 2567 aacttccctccacgaactgtgattgtcctgtgcagtataactgcgaagaaga 2622
RESULT 6
AAZ50875
ID AAZ50875 standard; cDNA: 2602 BP.
XX
AC AAZ50875;
XX
DT 31-MAY-2000 (first entry)
XX
DE Partial human GBS toxin receptor (HP55) cDNA.
XX
KW Human GBS toxin receptor; group B beta-haemolytic streptococci; HP55;
KW pathological vascularisation; cancer metastases; angiogenesis;
KW neovascularisation; reperfusion injury; scarring; keloid;
KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW endothelial cell proliferation; antibacterial; anticancer;
KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 58..1545
FT CDS /*tag= a
FT /product= "Partial human GBS toxin receptor"
XX
PN W0200005375-A1.
XX
PD 03-FEB-2000.

| | | | |
|----|------|---|------|
| OY | 1695 | gtgttaagatttgtaagcctgtgtatcaatgaatagtcacagttgcacataagcaaat | 1754 |
| Db | 1655 | gtataatattttaaaggccataatcatbtaaatatcatcagttgcgaataatataatg | 1714 |
| OY | 1755 | agcattcttaattatattatcaacccgtttgcctggaactcaatcaaggtcaacatct | 1814 |
| Db | 1715 | aacgtgttcaatatatgaataataatgtaagctggaactctactttagttccatact | 1774 |
| OY | 1815 | ggtcgaagtcagcaaccacacataaggggagttcatlta- tltataagaacataacc | 1873 |
| Db | 1775 | gcccgtcagtcoggcacaacagaaagtagcagaagctctgttatttttaaggccataacta | 1834 |
| OY | 1874 | aagsgtagagctgaataaagaccctctctataccttttgctaataaagrgygaataaatc | 1933 |
| Db | 1835 | agggaaagagcttgaataaagaccctctctataccttttgctaataaactagaataatlc | 1894 |
| OY | 1934 | tcaagctcttlttaaacatcgtctttgttgaacatcttcctcaaaaataatttgcatacgc | 1993 |
| Db | 1895 | tcaagtaactgataaacacacctgtgtgttgcataccttcctcataaa- ----atgtcaagc | 1948 |
| OY | 1994 | aatccctcgacaatgttagtttccaacttttagcctctccacgagagctggaagccatgtatc | 2053 |
| Db | 1949 | tctctcgacaactaagaccctcaacttttagcatctctgtgagcgccatccacatgtata | 2008 |
| OY | 2054 | atccagccttgcgaacttcatcttttggggaagacatgagccagctgcacatgtccacct | 2113 |
| Db | 2009 | attcgccttgcgaacttgcagcttggaggtgtgtgcccagacagctgcacagcatctcccc | 2068 |
| OY | 2114 | ctgagctcaaggagcaagtgcacgaacatt- ----agcagagatccaaagaccaggtcag | 2166 |
| Db | 2069 | ctgagcttcaaggtctcagaggtgcaccaggttatacagggagcagatccaaagccagagccag | 2128 |
| OY | 2167 | cgcgaagagccttggaaaggtatcttccct- -gggagctgttaagtgtgtgataagagccct | 2224 |
| Db | 2129 | tgttcagaccttgcgctgtgtccttccctccctgcaggggctatcaagtgtgataaagagccct | 2188 |
| OY | 2225 | gagcgaacagggagagcgagcatcacagatcatgtttccatcacacctc- tcccttccct | 2283 |
| Db | 2189 | gagtaggagcaagagcaggtgagatcccaactgtatgtgtcttgatatacattcccaacttccct | 2248 |
| OY | 2284 | tcccaagcacactgtgataattgctctgcacatgttaacctgcataaagaana | 2329 |
| Db | 2249 | tcccaagcacaggaatattgtgcttgcacatgttaacctgcataaagaana | 2294 |

| | | |
|----------|---|--------------------------|
| RESULT | 7 | |
| AAK94876 | | |
| ID | AAK94876 | standard; cDNA; 2712 BP. |
| XX | | |
| XX | | |
| AC | AAK94876; | |
| XX | | |
| DT | 06-NOV-2001 | (first entry) |
| XX | | |
| DE | Human full-length cDNA, SEQ ID NO: 4068. | |
| XX | | |
| XX | Human; full length cDNA; cDNA synthesis; oligo-capping; ss. | |
| XX | | |
| OS | Homo sapiens. | |
| XX | | |
| PN | EP130094-A2. | |
| XX | | |
| PD | 05-SEP-2001. | |
| XX | | |
| PF | 07-JUL-2000; 2000EP-0114089. | |
| XX | | |
| PR | 08-JUL-1999; 99JP-0194486. | |
| PR | 11-JAN-2000; 2000JP-0118774. | |
| PR | 02-MAY-2000; 2000JP-0183765. | |
| XX | | |
| XX | | |
| PA | (HELI-) HELIX RES INST. | |
| OT | Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y. | |

| | |
|----|---|
| PI | Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H; |
| XX | |
| DR | WPI, 2001-524255/58. |
| DR | P-PsDB: AAM93914. |
| XX | |
| PT | 830 Primers useful for synthesizing full length cDNA clones and their |
| PT | use in genetic manipulation - |
| XX | |
| PS | Claim 8, SEQ ID NO 4068; 1380bp + sequence listing; English. |
| XX | |
| CC | The invention relates to primers for synthesizing full length cDNA |
| CC | clones. 830 cDNA molecules encoding a human protein have been |
| CC | isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA |
| CC | molecules have been determined. Primers for synthesizing the full length |
| CC | cDNA are useful for clarifying the function of the protein encoded by |
| CC | the cDNA. The full length clones were obtained by construction of full |
| CC | length enriched cDNA libraries that were synthesised by the oligo-capping |
| CC | method. The primers enable the production of the full length cDNA easily |
| CC | without any special methods. The present sequence is a full length |
| CC | human cDNA of the invention. |
| CC | Note: The sequence data for this patent did not form part of the printed |
| CC | specification, but was obtained in CD-ROM format directly from EPO. |
| XX | |
| SQ | Sequence 2712 BP; 677 A; 619 C; 619 G; 797 T; 0 other; |

| Query Match | 51.3% | Score 1458.8 | DB 22 | Length 2712 |
|-----------------------|--------------|---|----------------|-------------|
| Best Local Similarity | 78.6% | Prod. No. 0 | | |
| Matches 1916 | Conservative | 0 | Mismatches 387 | Indels 135 |
| | | | Gaps | 9 |
| OY | 5 | gggggccccggttcctggcgccctccgtgagctcccttccgcggagcaagtttgcgc | 64 | |
| | | | | |
| Db | 6 | ggcccggggggctctggcgcttgctggcgccggcgctccctctctgcgaagttgcgataac | 65 | |
| OY | 65 | gtatccctcctgaaggcatcatgaagtcctccgggttcgacttaagcccgagcgacgcga | 124 | |
| | | | | |
| Db | 66 | ct-gcctacgtaagcgctcatatgagttctccgttcctcgtaacgttgcgcggaacgatgcga | 123 | |
| OY | 125 | ggagggcctcgagcccgacacgcctctcgaacgcgcgcgcgcgcgcgcgcgcgcgcaggt | 184 | |
| | | | | |
| Db | 124 | ggagggcctcgagcccgacacgcctctcctcccgggcgccccacagggcgccgcctccagtt | 183 | |
| OY | 185 | atgcgcctctgcctgccttaacactagaattttgtcctttttggttctcttcgtctcta | 244 | |
| | | | | |
| Db | 184 | gtgcgcctctgcctgttaacactagaacttggcctttttgtttcttcattgtta | 243 | |
| OY | 245 | ttcattaggggttgaaatctcgagcggttgcactagtggacttggttgatttaacaacacgcgc | 304 | |
| | | | | |
| Db | 244 | tgcatctaggttgaaatctcgaggtgttgctttagtgatattgaaatttaacaacacttc | 303 | |
| OY | 305 | caaaagataagaaagctctcctcgaagtcgagcgagcatctgcctccatacaaaagtctctca | 364 | |
| | | | | |
| Db | 304 | agaagataataagaactctcaagggcggtgcagagagatctgtcccatataaaagtctacaa | 363 | |
| OY | 365 | caaccaaaccgggttaaaaaagtaccgggttgatgagtcagaaactcaagatgattcttcgagtc | 424 | |
| | | | | |
| Db | 364 | taattcaaaacgggttaagaagttccaaatggatgagcagaaaactcaagatgattcttcggttc | 423 | |
| OY | 425 | ttttctcatatgctacatcatcaacaacaattcccttgaaagatatgttcgcacagaagtg | 484 | |
| | | | | |
| Db | 424 | ctttttatgtgtctacatcatcaacaacaagattcccttgaaagatatgttcgcagcaaaatagg | 483 | |
| OY | 485 | ggggaagctgtgtcaggatctcggagatctttgtctacagatatcctcacctgttcaactcc | 544 | |
| | | | | |
| Db | 484 | gggggaaaatgtctcagtagattttggatctcttgcactctgtccctcaaccctgttcaactcc | 543 | |
| OY | 545 | ctctgcctcagatctcggagtcggaagccctgtttgcaatcaggcgcaacagagaaggtcagg | 604 | |
| | | | | |
| Db | 544 | catgtctcgaaatttaggaagttggaacccactcatgttactcaagagcaactagaagaactagg | 603 | |
| OY | 605 | agaaggtgtcacaataccagccaatgcatagtcagatgtgtctcatatgggtcccccctctga | 664 | |
| | | | | |
| Db | 604 | agaaggttttcaattccagccaatgcatagtcagatgtgtgtctctttaggtcccccctctga | 663 | |

| | | | |
|----|------|--|------|
| QY | 665 | aagaagcaagctcttgatgtatttcataatgcaggagcaaaacttggagcaagtattctct | 724 |
| Db | 664 | aagaagcaaaattctctttagcatcttcataatcaggagacacagctttggacagataattctct | 723 |
| QY | 725 | tcctctttctggaaataatttgcctactatbtaattggaacttatgtctctcttctttg | 784 |
| Db | 724 | tcctcttctcggaaataatttgcctacataagaattggaacttatgtctctctcttctttg | 783 |
| QY | 785 | catgtttggaatcctcgtgttatttlatvgatctccttagttagtgaataacagaagac | 844 |
| Db | 784 | tactatttggaaataatttgttcttcttcttggatcgtttagttagtgcacacaacaataa | 843 |
| QY | 845 | tcaacaagaacaatcactccctcctatgaaaaaggatataattcttcataatnaaaat---- | 899 |
| Db | 844 | acacaaagaagaattcccatatattgaaaaggataacattcttcataatnaaatacagc | 903 |
| QY | 900 | ----- | 899 |
| Db | 904 | tggagttgcagttagcatgatctttggtcctaactgcagctccctctgcctccggtttcaagc | 963 |
| QY | 900 | -----caagctctcttcacagaagtcaat | 922 |
| Db | 964 | catcctccacactcagccttacaagaagttagttaggcagaacagcttcttcacaagaagtcaat | 1023 |
| QY | 923 | gcgcgtgataactatgctgaataatcactgcgaactttgggctatgttcgtttgcacatttc | 982 |
| Db | 1024 | gcgcgttggtaacctattttaaataacctgcgaactttgggctatgcgttagttagcacatttc | 1083 |
| QY | 983 | ttacaactgcgaacttttaactcttgtttgcacttatgtgcctacttatcagaaggagctct | 1042 |
| Db | 1084 | ttacaacgcgacttttaactcttcatcttagcatatgtcgaacttatcagaaggagactct | 1143 |
| QY | 1043 | aaggttcaataatcacaagaagaattggttttatctgcagtcgaccttattaggtttgtgt | 1102 |
| Db | 1144 | aagtttcoaattgttcaagaagaattggttttatctgcacttatgtcttattaggtcttggtt | 1203 |
| QY | 1103 | atgtatgatactcgttcgggttcaagctgcgcgaacaattaaaggcagaatgaaattttcaac | 1162 |
| Db | 1204 | atgtatgatcctcgttcgttcaagctgcgcgaacaattaaaggcagaatgaaattttcaac | 1263 |
| QY | 1163 | tcgttgggttcggaagagtttttgcgcttataggagatgattggaactcgtgatatctcgtgt | 1222 |
| Db | 1264 | ttctatggttcgcgcgaatcttcttagaccttataggaaagatgtgacctgcgaattctcgtgt | 1323 |
| QY | 1223 | ttgcgcgacgaatttataagcgtgtgatatattcctctgcgttcgttgcattcccaacatataac | 1282 |
| Db | 1324 | agctgcgtgcctcaattgctgcgtgatatctcttgcgcgtgttcctccaataataataac | 1383 |
| QY | 1283 | aaaccttggagagcttttgcctcttcctcgtgatttagcatcaacaactcgtgaacttgcctctc | 1342 |
| Db | 1384 | aaaccttggagagcttttgcctcttcctcgtgatttagcatcaacaacaactcgtgatatgtccttc | 1443 |
| QY | 1343 | gtatgcgcgatatctcctcgtggcactcaaaatccctttgcgaactattcctcttgaatgatttg | 1402 |
| Db | 1444 | gtatgcgtgcgatacctcctcgtggcactcaaaatcatttgcgaactattccagaatgtgttg | 1503 |
| QY | 1403 | gcccatatgcgaagaatttcaacccctgaagaacatatttggagaattgcaaacctgtttt | 1462 |
| Db | 1504 | gcccgctaatgtcacaagaagctcgcgaacccctgataacaactgtttgagaaatggcaaacccgtgt | 1563 |
| QY | 1463 | ctgcatacgtcgtcgtatacaatgtatctttgttgcactttctcttcacaacattcgcgcaag | 1522 |
| Db | 1564 | ctatattgtcgtcgtcattaaattgttttggcgaatttcttctaacattcgcgcaag | 1623 |
| QY | 1523 | tgaagtgcaaaactgtggcactcaagtgaatcaacaaggacacagaagaactgaaaggaacata | 1582 |
| Db | 1624 | tgaagtgcaaaactgtggcctcctaactgatcaacaatgcagacagacacactgaaaggaacata | 1683 |
| QY | 1583 | ataatactgtcctccttataatgatatactttgttatcatcagttaacctcaaaagcgctttgat | 1642 |
| Db | 1684 | ataatactgtcctccttataatgtatatttataattatcatcagttaacctcaaaagcgctt--- | 1739 |

| | | | |
|----------|---|--|------|
| QY | 1643 | atcttaatgtgtgaagcaactatatacaagaataaatttgcactgagaaaattggtgttaa | 1702 |
| DY | 1740 | -ctgtatctgtgaagcaatctcattgt------ctttttaattgtactgttattaga | 1789 |
| QY | 1703 | ttctgaagccttgtaatactgaagtgtcaactggttgcataaagaacttagctattc | 1762 |
| DB | 1790 | tttttaaggccataatactcgtgaataatacctcgttgcgcgataataaataagacgtgt | 1849 |
| QY | 1763 | ttaattatattaaacccgtttgcgtgaacttaacaattcgaaggttcacaratctgcgtcaa | 1822 |
| DB | 1850 | ttcaattatgaataataatgaagcagactcttcaacttagtttcacataaccgtcgtcta | 1909 |
| QY | 1823 | gtccaggtcaacccacaatatagggagttctattta-ttataagaacctacttaagaagatg | 1881 |
| DB | 1910 | gtccgggtcaacatgaagtaaggaagctctctgttattttttaagggtcacattacaaaggaatg | 1969 |
| QY | 1882 | agcttaaatagaacctctcatcacttctgttgaattaaagggtgataataattctcagctct | 1941 |
| DB | 1970 | agctgtaaaaagagccctccgtgatacactcttgcgttaattaaactagaatgataattctcagttac | 2029 |
| QY | 1942 | tgttaacacatcgtttttgtgtacaccccttcctcaaaaaaattattgttcataagcaatccctg | 2001 |
| DB | 2030 | tgtataacacccctgtgtgtgttcacattgtctcataaaa-----atgtcagctctctctg | 2083 |
| QY | 2002 | acaatgtagttctccaactttgaacctctccacggagactgtgcacgcactgtatcaatcagcc | 2061 |
| DB | 2084 | acaacttagacccctcaaacctttagcattctctgtgtgagctgtgcacccactgtataatttcggcc | 2143 |
| QY | 2062 | tggcaacttcacattgtgggaagacatgtccacggagcgtgcacacatgtccctctctgtgcttc | 2121 |
| DB | 2144 | tggcaactcggactgtggggaggtgtgtccacggagcagctgtgcgaagcactcctcctctgtcttc | 2203 |
| QY | 2122 | agggaacagttgcccaagcaact-----aggcagacatcccaagccacgggtcgaagcgccaag | 2174 |
| DB | 2204 | agggtcagagatgcccagcgcttatcagaaggcagacatcccaagccagagccagtcgtctgact | 2263 |
| QY | 2175 | ctttgaagcgtatcttctccccc--gggagctgttaatgtgtgtgataagccctgagccaac | 2232 |
| DB | 2264 | cttcgcgtctgtgccttctctcctctgaggggctataccaatgtgtgataaagccctggaatgagc | 2323 |
| QY | 2233 | agggaacagcgagatccacagatcatgtttcattcattgcacctc--tcccttccctccacaga | 2291 |
| DB | 2324 | aagagcagatgagatccacactgcgtcatgtgtctgtatatacctcaaacctttcccttccacaga | 2383 |
| QY | 2292 | cacttgagatgtcctgagcatgttlaacctgtcaaaaagaaa | 2329 |
| DB | 2384 | cagagggaattatgtctgcatgtcaacccgtcaaaaagaaa | 2421 |
| RESULT 8 | | | |
| AAZ50880 | AAZ50880 standard; cDNA; 1485 BP. | | |
| NC | AAZ50880; | | |
| XX | 31-MAY-2000 (first entry) | | |
| DT | Human/Sheep consensus GBS toxin receptor cDNA sequence-1. | | |
| XX | Human GBS toxin receptor; group B beta-haemolytic streptococci; Hp59; | | |
| KW | pathological vascularisation; cancer metastases; anglogenesis; | | |
| KW | neovascularisation; reperfusion injury; scarring; keloid; | | |
| KW | chronic inflammation; rheumatoid arthritis; psoriasis; neural injury; | | |
| KW | endothelial cell proliferation; antibacterial; antitumour; | | |
| KW | anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss. | | |
| XX | Homo sapiens. | | |
| OS | Ovis sp. | | |
| EH | Key | | |
| ET | mat_peptide | | |
| ET | 1..1485 | | |
| ET | /*tag= a | | |
| ET | /product= "Human/sheep consensus GBS toxin receptor | | |

protein sequence"
FT /transl_except- (pos:4..6; aa:Xaa)
FT /transl_except- (pos:7..9; aa:Xaa)
FT /transl_except- (pos:16..18; aa:Xaa)
FT /transl_except- (pos:22..24; aa:Xaa)
FT /transl_except- (pos:28..30; aa:Xaa)
FT /transl_except- (pos:31..33; aa:Xaa)
FT /transl_except- (pos:34..36; aa:Xaa)
FT /transl_except- (pos:46..48; aa:Xaa)
FT /transl_except- (pos:49..51; aa:Xaa)
FT /transl_except- (pos:58..60; aa:Xaa)
FT /transl_except- (pos:61..63; aa:Xaa)
FT /transl_except- (pos:64..66; aa:Xaa)
FT /transl_except- (pos:67..69; aa:Xaa)
FT /transl_except- (pos:70..72; aa:Xaa)
FT /transl_except- (pos:73..75; aa:Xaa)
FT /transl_except- (pos:79..81; aa:Xaa)
FT /transl_except- (pos:85..87; aa:Xaa)
FT /transl_except- (pos:91..93; aa:Xaa)
FT /transl_except- (pos:100..102; aa:Xaa)
FT /transl_except- (pos:124..126; aa:Xaa)
FT /transl_except- (pos:130..132; aa:Xaa)
FT /transl_except- (pos:136..138; aa:Xaa)
FT /transl_except- (pos:154..156; aa:Xaa)
FT /transl_except- (pos:157..159; aa:Xaa)
FT /transl_except- (pos:163..165; aa:Xaa)
FT /transl_except- (pos:169..171; aa:Xaa)
FT /transl_except- (pos:181..183; aa:Xaa)
FT /transl_except- (pos:187..189; aa:Xaa)
FT /transl_except- (pos:190..192; aa:Xaa)
FT /transl_except- (pos:196..198; aa:Xaa)
FT /transl_except- (pos:202..204; aa:Xaa)
FT /transl_except- (pos:211..213; aa:Xaa)
FT /transl_except- (pos:220..222; aa:Xaa)
FT /transl_except- (pos:223..225; aa:Xaa)
FT /transl_except- (pos:235..237; aa:Xaa)
FT /transl_except- (pos:241..243; aa:Xaa)
FT /transl_except- (pos:244..246; aa:Xaa)
FT /transl_except- (pos:250..252; aa:Xaa)
FT /transl_except- (pos:277..279; aa:Xaa)
FT /transl_except- (pos:280..282; aa:Xaa)
FT /transl_except- (pos:283..285; aa:Xaa)
FT /transl_except- (pos:295..297; aa:Xaa)
FT /transl_except- (pos:304..306; aa:Xaa)
FT /transl_except- (pos:337..339; aa:Xaa)
FT /transl_except- (pos:340..342; aa:Xaa)
FT /transl_except- (pos:346..348; aa:Xaa)
FT /transl_except- (pos:367..369; aa:Xaa)
FT /transl_except- (pos:394..396; aa:Xaa)
FT /transl_except- (pos:397..399; aa:Xaa)
FT /transl_except- (pos:406..408; aa:Xaa)
FT /transl_except- (pos:409..411; aa:Xaa)
FT /transl_except- (pos:412..414; aa:Xaa)
FT /transl_except- (pos:421..423; aa:Xaa)
FT /transl_except- (pos:430..432; aa:Xaa)
FT /transl_except- (pos:433..435; aa:Xaa)
FT /transl_except- (pos:436..438; aa:Xaa)
FT /transl_except- (pos:442..444; aa:Xaa)
FT /transl_except- (pos:445..447; aa:Xaa)
FT /transl_except- (pos:463..465; aa:Xaa)
FT /transl_except- (pos:475..477; aa:Xaa)
FT /transl_except- (pos:481..483; aa:Xaa)
FT /transl_except- (pos:487..489; aa:Xaa)
FT /transl_except- (pos:490..492; aa:Xaa)
FT /transl_except- (pos:493..495; aa:Xaa)
FT /transl_except- (pos:496..498; aa:Xaa)
FT /transl_except- (pos:502..504; aa:Xaa)
FT /transl_except- (pos:514..516; aa:Xaa)
FT /transl_except- (pos:529..531; aa:Xaa)
FT /transl_except- (pos:535..537; aa:Xaa)
FT /transl_except- (pos:562..564; aa:Xaa)
FT /transl_except- (pos:589..591; aa:Xaa)
FT /transl_except- (pos:595..597; aa:Xaa)

FT /transl_except- (pos:598..600; aa:Xaa)
FT /transl_except- (pos:604..606; aa:Xaa)
FT /transl_except- (pos:619..621; aa:Xaa)
FT /transl_except- (pos:634..636; aa:Xaa)
FT /transl_except- (pos:655..657; aa:Xaa)
FT /transl_except- (pos:691..693; aa:Xaa)
FT /transl_except- (pos:694..696; aa:Xaa)
FT /transl_except- (pos:700..702; aa:Xaa)
FT /transl_except- (pos:703..705; aa:Xaa)
FT /transl_except- (pos:706..708; aa:Xaa)
FT /transl_except- (pos:712..714; aa:Xaa)
FT /transl_except- (pos:715..717; aa:Xaa)
FT /transl_except- (pos:724..726; aa:Xaa)
FT /transl_except- (pos:727..729; aa:Xaa)
FT /transl_except- (pos:736..738; aa:Xaa)
FT /transl_except- (pos:748..750; aa:Xaa)
FT /transl_except- (pos:757..759; aa:Xaa)
FT /transl_except- (pos:760..762; aa:Xaa)
FT /transl_except- (pos:769..771; aa:Xaa)
FT /transl_except- (pos:772..774; aa:Xaa)
FT /transl_except- (pos:775..777; aa:Xaa)
FT /transl_except- (pos:778..780; aa:Xaa)
FT /transl_except- (pos:790..792; aa:Xaa)
FT /transl_except- (pos:793..795; aa:Xaa)
FT /transl_except- (pos:811..813; aa:Xaa)
FT /transl_except- (pos:820..822; aa:Xaa)
FT /transl_except- (pos:847..849; aa:Xaa)
FT /transl_except- (pos:850..852; aa:Xaa)
FT /transl_except- (pos:853..855; aa:Xaa)
FT /transl_except- (pos:856..858; aa:Xaa)
FT /transl_except- (pos:862..864; aa:Xaa)
FT /transl_except- (pos:880..882; aa:Xaa)
FT /transl_except- (pos:883..885; aa:Xaa)
FT /transl_except- (pos:892..894; aa:Xaa)
FT /transl_except- (pos:922..924; aa:Xaa)
FT /transl_except- (pos:928..930; aa:Xaa)
FT /transl_except- (pos:943..945; aa:Xaa)
FT /transl_except- (pos:952..954; aa:Xaa)
FT /transl_except- (pos:955..957; aa:Xaa)
FT /transl_except- (pos:970..972; aa:Xaa)
FT /transl_except- (pos:994..996; aa:Xaa)
FT /transl_except- (pos:997..999; aa:Xaa)
FT /transl_except- (pos:1009..1011; aa:Xaa)
FT /transl_except- (pos:1012..1014; aa:Xaa)
FT /transl_except- (pos:1033..1035; aa:Xaa)
FT /transl_except- (pos:1063..1065; aa:Xaa)
FT /transl_except- (pos:1081..1083; aa:Xaa)
FT /transl_except- (pos:1084..1086; aa:Xaa)
FT /transl_except- (pos:1090..1092; aa:Xaa)
FT /transl_except- (pos:1096..1098; aa:Xaa)
FT /transl_except- (pos:1111..1113; aa:Xaa)
FT /transl_except- (pos:1126..1128; aa:Xaa)
FT /transl_except- (pos:1129..1131; aa:Xaa)
FT /transl_except- (pos:1138..1140; aa:Xaa)
FT /transl_except- (pos:1141..1143; aa:Xaa)
FT /transl_except- (pos:1144..1146; aa:Xaa)
FT /transl_except- (pos:1147..1149; aa:Xaa)
FT /transl_except- (pos:1150..1152; aa:Xaa)
FT /transl_except- (pos:1153..1155; aa:Xaa)
FT /transl_except- (pos:1168..1170; aa:Xaa)
FT /transl_except- (pos:1174..1176; aa:Xaa)
FT /transl_except- (pos:1180..1182; aa:Xaa)
FT /transl_except- (pos:1189..1191; aa:Xaa)
FT /transl_except- (pos:1201..1203; aa:Xaa)
FT /transl_except- (pos:1246..1248; aa:Xaa)
FT /transl_except- (pos:1270..1272; aa:Xaa)
FT /transl_except- (pos:1291..1293; aa:Xaa)
FT /transl_except- (pos:1306..1308; aa:Xaa)
FT /transl_except- (pos:1315..1317; aa:Xaa)
FT /transl_except- (pos:1324..1326; aa:Xaa)
FT /transl_except- (pos:1330..1332; aa:Xaa)
FT /transl_except- (pos:1333..1335; aa:Xaa)
FT /transl_except- (pos:1339..1341; aa:Xaa)

| | | |
|----|-----------------------|--------------------------|
| FT | /transl_except= | (pos:1150..1152; aa:Xaa) |
| FT | /transl_except= | (pos:1153..1155; aa:Xaa) |
| FT | /transl_except= | (pos:1158..1170; aa:Xaa) |
| FT | /transl_except= | (pos:1174..1176; aa:Xaa) |
| FT | /transl_except= | (pos:1180..1182; aa:Xaa) |
| FT | /transl_except= | (pos:1189..1191; aa:Xaa) |
| FT | /transl_except= | (pos:1201..1203; aa:Xaa) |
| FT | /transl_except= | (pos:1246..1248; aa:Xaa) |
| FT | /transl_except= | (pos:1270..1272; aa:Xaa) |
| FT | /transl_except= | (pos:1291..1293; aa:Xaa) |
| FT | /transl_except= | (pos:1306..1308; aa:Xaa) |
| FT | /transl_except= | (pos:1315..1317; aa:Xaa) |
| FT | /transl_except= | (pos:1324..1326; aa:Xaa) |
| FT | /transl_except= | (pos:1330..1332; aa:Xaa) |
| FT | /transl_except= | (pos:1333..1335; aa:Xaa) |
| FT | /transl_except= | (pos:1339..1341; aa:Xaa) |
| FT | /transl_except= | (pos:1348..1350; aa:Xaa) |
| FT | /transl_except= | (pos:1357..1359; aa:Xaa) |
| FT | /transl_except= | (pos:1372..1374; aa:Xaa) |
| FT | /transl_except= | (pos:1375..1377; aa:Xaa) |
| FT | /transl_except= | (pos:1381..1383; aa:Xaa) |
| FT | /transl_except= | (pos:1384..1386; aa:Xaa) |
| FT | /transl_except= | (pos:1396..1398; aa:Xaa) |
| FT | /transl_except= | (pos:1402..1404; aa:Xaa) |
| FT | /transl_except= | (pos:1420..1422; aa:Xaa) |
| FT | /transl_except= | (pos:1444..1446; aa:Xaa) |
| FT | /transl_except= | (pos:1456..1458; aa:Xaa) |
| FT | /transl_except= | (pos:1459..1461; aa:Xaa) |
| FT | /transl_except= | (pos:1462..1464; aa:Xaa) |
| FT | /transl_except= | (pos:1471..1473; aa:Xaa) |
| FT | /transl_except= | (pos:1483..1485; aa:Xaa) |
| FT | /note="Xaa = unknown" | |

PN WO200005375-A1.
 XX
 PD 03-FEB-2000.
 XX
 PF 22-JUL-1999; 99WO-US16676.
 XX
 PR 22-JUL-1998; 98US-0093843.
 XX
 PA (UYVA-) UNIV VANDERBILT.
 XX
 PI Høllergvist CG, Fu C;
 XX
 DR WPI; 2000-205377/18.
 XX P-PSDB; AA45090.
 XX

| Query Match | 45.98% | Score 1305 | DB 21 | Length 1485 |
|-----------------------|---|------------------|----------------|-------------|
| Best Local Similarity | 87.9% | Pred. No. 5e-293 | | |
| Matches 1305 | Conservative | 0 | Mismatches 180 | Indels 0 |
| | | | | Gaps |
| QY 84 | atgaagtcgccggtttcggacttagcccgcgagcgagcgagggctcgaccgcaca | 1434 | | |
| Db 1 | atgaagtlncgcggtltnnngaactnccngcngacngagcgagagngcncgagaccacm | 60 | | |
| QY 144 | ccgcctcgcgcagcgccgccccgcggcggaacccgtccagatctctgcctcgctac | 2030 | | |
| Db 61 | ccnctnccnccngmcgccccnccgagcggaancgcgtcccaatnngctctgcctgttac | 1200 | | |
| QY 204 | aacctagcattttgtccctttttgtgtctctgtctctatcattacagggatgactcg | 2630 | | |
| Db 121 | aactnagaanttttgcctttttgtgtctctctctctnttntatnattagngtgaactcg | 1800 | | |
| QY 264 | agcgttgcactagtgagacatggtgattcaacaacaacgcccagaagataagaaagctcc | 3220 | | |
| Db 181 | agngttcgmatagtgagaaatggtatgtatcaanaacaactnnnaagtaataagacnctcc | 2400 | | |
| QY 324 | tagagatgtgtagagcgacttcgtcccccataaagaatttcttcaacaaccaaaggttaaaag | 3530 | | |
| Db 241 | nanngtvttnaagagcatcttcgctcccataaaagtttctnanaancacaacgggttaaaag | 3000 | | |

| | | | |
|----|------|---|------|
| QY | 384 | taccgvtggaibcgaagaactcgaagtatctctcgatcttcttcattgctacatc | 443 |
| Db | 301 | taccmhtvggatvgagaactcgaagtatctctcgatcttcttcattgctacatc | 360 |
| QY | 444 | atcaacaatatctctcgagagatagtttgcagcagaagvtggggggaagcgtgtgcagta | 503 |
| Db | 361 | atcaacaatatctctcgagagatagtttgcagcagaanmggggggaamtgtgtcagga | 420 |
| QY | 504 | ttcgagatcttgtcaagcatctctcaaccctgttcaactccctctgcgtcagaatttcga | 563 |
| Db | 421 | ttnggatcmttgmagmcgtmctntcaccctgttcaactccctcmtngctcgagattngga | 480 |
| QY | 564 | gtcggaagcccttgttgcactcagggcactagaaggctagagaggtgtcacaatacca | 623 |
| Db | 481 | gtngagmccmctnttgmactcagacactagaagmctagaagaggtgttmaacttcca | 540 |
| QY | 624 | gccatgcactgcagvtggtcttcaatgggcctccctcttgaagaagcgaagcttcgaat | 683 |
| Db | 541 | gccatgcactgcagvtggtcttcaatgggcctccctcttgaagaagcgaagcttcgaat | 600 |
| QY | 684 | attcatactgagagagacaacaacttggagacagtagtttctctctctctcttcgtagtaatt | 743 |
| Db | 601 | attcmctatgcagagagacaancttggagacagtagtttctctctctcttcgtagtaatt | 660 |
| QY | 744 | tgtactataatgatvtgacttatgtctctatcttcttggcatgtgttgaatlcactg | 803 |
| Db | 661 | tgtactataatgatvtgacttatgtctctatcttcttggmanhttvgaatmhtntvg | 720 |
| QY | 804 | tttatattatgatctgtcttagttatgatacaccagaactcacaagacaatctacc | 863 |
| Db | 721 | ttnttnttngatctgtntttagttatgatacaccanaamcaagaanaatmctmcm | 780 |
| QY | 864 | tatgaaagagagataattcttcttcaatlaaaaaatcagctctcttcacagaagtcagt | 923 |
| Db | 781 | tatgaaagagagatanaattcttcttcaatlaanaaalcagctntcttcacagaagtcagt | 840 |
| QY | 924 | cctgtgatacctaactgtcgaatacactgcacttgggcattgtgcgttgcacatttct | 983 |
| Db | 841 | cctgtgatacctaactmhtnaaatcmctgcacttgggcattngtctgcacatttct | 900 |
| QY | 984 | tacaaactggaacttttattacttgttgacttttgcacttatcaatgaagaagtctta | 1043 |
| Db | 901 | tacaaactggaacttttattacttnttgaacttatgtccactatanaagagaamtctta | 960 |
| QY | 1044 | aggttcaatatcaagaagaatvtgggttttatactgcagctccactatnaatgaagttgtgta | 1103 |
| Db | 961 | aggttcaatatcaagaagaatvtgggttttatactcncanmhtcattatgaagmtgtgta | 1020 |
| QY | 1104 | tgtatgatctgtctgggtlcaagctgtcgaactttaaaggagcaagatlgaaatttccaact | 1163 |
| Db | 1021 | tgtatgatctgtctgggtlcaagctgtcgaactttaaaggagcaanaatlgaaatttccaact | 1080 |
| QY | 1164 | ctgtgaggttcgaagaagtttttaagcctatagaatgaattggacgtcgatctctgtt | 1222 |
| Db | 1081 | ntmngnctgcgaagaantlcttaagcctatagaatgaatgtgagactvgcmhtatctctgttn | 1140 |
| QY | 1224 | ggccagagatttlaaggctgtgatatctctctgcgttgcattcccttcaacatatacaaa | 1283 |
| Db | 1141 | ggcmngmntnaatnagctgtgatatctnttvgcmgttgntcctcaacatatacaaa | 1200 |
| QY | 1284 | acccttggagagcttvtgcctctctcgatttagcaacaacactcgtgacatgtgcctctg | 1343 |
| Db | 1201 | acmcttggagagcttvtgcctctctcgatttagcatcaacactcgtganaattgtcctctg | 1260 |
| QY | 1344 | tatgctgttatctcctcgtggcatcacaanaactttgccaactatctcgtgaatgttgg | 1403 |
| Db | 1261 | tatgctgttatctcctcgtggcatcacaanaactttgccaactatctcgtgaatgttgg | 1320 |
| QY | 1404 | cccatcatctgcagaagaatttcaacctgcagaagaacattggaggaatgtgcaaatcttctc | 1463 |
| Db | 1321 | cccatcatctgcagaagaatttcaacctgcagaagaacacmcttggaggaatgtgcaaaingntctc | 1380 |
| QY | 1464 | tgcactgcgtctgtactaagaatgtattgtgtgcatttcttcaacatattgcgcaaggt | 1523 |

Db 961 aggtcaatgttcaagaagaatgggttttatcttcaatgtccttatttagtccttggtta 1020
QY 1104 tgaatgacccgttcgagcaagctgactgacaaattaaaggcaagatggaattttcaact 1163
Db 1021 tgaatgacccgttcgagcaagctgactgacaaattaaaggcaagatggaattttcaact 1080
QY 1164 ctgttggttcgaagaagtttttagccttaagagbatttgaccctgacatcttcctggtt 1223
Db 1081 ttatgtgtcgcagaatttttagccttaagagbatttgaccctgacatcttcctggtta 1140
QY 1224 gccgagagattatagctgtgattattcctctgtctgttgcattcctcaacatacaaca 1283
Db 1141 gctgctggtcttcattgctgtgattattcctctgtctgttgcattcctcaacatacaaca 1200
QY 1284 accctggagagctttgtctctctccttgattagacacacatcttgagacattgtccttcg 1343
Db 1201 acactggagagctttgtctctctccttgattagacacacatcttgagacattgtccttcg 1260
QY 1344 taagtgtattctccttgagcaacaataccttgcacattccttgaaatgattgg 1403
Db 1261 tatgtgtatccctccttgagcaacaataccttgcacattccttgaaatgattgg 1320
QY 1404 cccatcatcgcagaagaagcttaacccttgagaaactatggaagaatggcaactgttttc 1463
Db 1321 cccgtcatcgttaaaagctcgaacccctgatacaactgttggagaatggcaaacctgttttc 1380
QY 1464 tgcctcgtctgctatcaatgatttggctgacatttcttcacacattcgcgcaaaagt 1523
Db 1381 tatattgtctgtctatcaatgatttggctgacatttcttcacacattcgcgcaaaagt 1440
QY 1524 gaagtgcacaaactgggcacatgattgataccagaagaacagaactga 1571
Db 1441 gaagtgcacaaactgggcacatgattgataccagaagaacagaactga 1488

RESULT 11
AAH79234
ID AAH79234 standard; cDNA; 2670 BP.
XX
AC AAH79234;
XX
DT 20-NOV-2001 (first entry)
XX
DE Human sodium dependent phosphate co-transfer protein 35 cDNA.
XX
KW Human; sodium dependent phosphate co-transfer protein 35;
KW hypophosphaturia; hypercalcaemia; hypophosphataemic rickets;
KW nephritis; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN CN1298882-A.
XX
PD 13-JUN-2001.
XX
PF 06-DEC-1999; 99CN-0124217.
XX
PR 06-DEC-1999; 99CN-0124217.
XX
PA (SHEN-) SHENGYUAN GENE DEV CO LTD SHANGHAI.
XX
PI Mao Y, Xie Y;
XX
DR WP1; 2001-503367/56.
XX
DR P-PSDB; AAG65238.
XX
PS
PS Human Na-dependent phosphate cotransporter 35 and its coding sequence -
PI Claim 6; Page 19-20(Disclosure); 28pp; Chinese.
XX
CC The present invention provides the protein and coding sequences of human
CC sodium dependent phosphate co-transfer protein 35. The sequences can be
CC used in the treatment of hypophosphaturia, hypercalcaemia,

CC hypophosphataemic rickets and nephritis. The present sequence is the
CC coding sequence of the invention.
XX
SQ Sequence 2670 BP; 706 A; 550 C; 552 G; 862 T; 0 other;

Query Match 41.1%; Score 1169.8; DB 22; Length 2670;
Best Local Similarity 82.6%; Pred. No. 1.5e-261;
Matches 1434; Conservative 0; Mismatches 272; Indels 31; Gaps 7;

QY 604 gaggaggtgtacaaatccacagcatgcatgctgactgcttctcattggtcctcccttg 663
Db 1 gagaggggtgtacaaatccacagcatgcatgctgactgcttctcattggtcctcccttg 60
QY 664 aaagaagaagcttcagatattcacaatgaggaagacaaacttggagagtagttcttc 723
Db 61 aaagaagaagcttcagatattcacaatgaggaagacaaacttggagagtagttcttc 120
QY 724 ttctctcttcgagtaattgtctactataatgaattggaactatgtctctattcttg 783
Db 121 ttctctcttcgagtaattgtctactataatgaattggaactatgtctctattcttg 180
QY 784 gcaattgtggaatcactcgtttattttaaagacatgcttaagttagtatacaccagaa 843
Db 181 gtaattgtggaatcactcgtttattttaaagacatgcttaagttagtatacaccagaa 240
QY 844 ctcaagaagacatcccccataaagaagatattcttctcaatcaataaaatacagc 903
Db 241 aacacaagaagatccccaataaagaagatattcttctcaatcaataaaatacagc 300
QY 904 tctctcagaagaatcagtgccgttgatgatacctatgctgaatcactgccaacttgggcta 963
Db 301 ttctctcagaagaatcagtgccgttgatgatacctatgataccttgaatccacttgggcta 360
QY 964 ttgtcgttgcacattttcttcaacacgtgacttttattactgtttagacttattgcta 1023
Db 361 ttgtcgttgcacattttcttcaacacgtgacttttattactgtttagacttattgcta 420
QY 1024 cttaacgaagaagatcccaatgaagttcaatatacaagaagaatgggttttcttcagctc 1083
Db 421 cttaacgaagaagatcccaatgaagttcaatatacaagaagaatgggttttcttcagctc 480
QY 1084 cttaattagttgtgttattgtatgtatccctgctgggttcaagctgctgaacaattaaagg 1143
Db 481 cttaattaggtcctctgttattgtatgtatccctgctgggttcaagctgctgaacaattaaagg 540
QY 1144 caaagtgaatttttcaacatcgtgggttcgaagaagtttttagccttataagagatgtg 1203
Db 541 caaagtgaatttttcaacatcgtgggttcgaagaagtttttagccttataagagatgtg 600
QY 1204 gacctgcgatatccctgtgttcgagagattatagcctgttattctcttggttg 1263
Db 601 gacctgcgatatccctgtgttcgagagattatagcctgttattctcttggttg 660
QY 1264 catcctaacaatactcaacaacacctgggaagcttttgcctctccttgatattagaacaac 1323
Db 661 catcctaacaatactcaacaacacctgggaagcttttgcctctccttgatattagaacaac 720
QY 1324 atctggacattgtctcctctgtatgtgtattctcctgggacatacaaatccttggca 1383
Db 721 atctggacattgtctcctctgtatgtgtattctcctgggacatacaaatccttggca 780
QY 1384 ctattcttggaatgattgggcccatactgacagaagcttcaaccctggagaacaactatg 1443
Db 781 ctattcttggaatgattgggcccatactgacagaagcttcaaccctggagaacaactatg 840
QY 1444 ggaatgcaaacatgtttctgcatcgtctgtctatcaatgattatgtgtgccaatttct 1503
Db 841 ggaatgcaaacatgtttctgcatcgtctgtctatcaatgattatgtgtgccaatttct 900
QY 1504 tcaacatactgcgcaaaagtgtgaagtcaaaactgagcactcagatgatacacaagaagaca 1563
Db 901 tcaacatactgcgcaaaagtgtgaagtcaaaactgagcactcagatgatacacaagaagaca 960


```
Db 1795 TCCTGGGCATCACAAATATACATTTGACCATCTATTCAGGAATGTTGGGCCCGCATTCGTCTA 1736
QY 1417 gaagcttacccttggaagaacattggagaatggcaaacctgtttcttgatgcctgctg 1476
Db 1735 AAGGCTTGACCCCTGATATACACTGTGTGGAGAAATGGCAAAACGGTGTCTATATTTGCTGCTG 1676
QY 1477 ctatcaatgtatttggctgacatttctctcaca--ctatctgcacaaagtgaaagtgcaaa 1534
Db 1675 CTATTTATATGTTTTTGGGCACTTTTCTTTAACTATTTTGCACAAAGGTGATACAA 1616
QY 1535 ctgggcatcagtgtatccaaagagacagaactgaaggaaaccaataaataatcctgtc 1594
Db 1615 CTGGGCTCTCAATGATCACATGACACACACACTGAAAGAACCAATAAATATCTCTGCC 1556
QY 1595 tctattaatgtatcttggttatcatgtatacctaaagtccttgatattttaaagtgt 1654
Db 1555 TCTATTATATATTTTATTTATTTATCATGTATACCTCAAAAGTCCCTT-----CTGATTTGTTG 1501
QY 1655 aagcaatctatatacaagaataaattgtactagaaaaaattgtgttagatttgaagctt 1714
Db 1500 AAGCATTTCTATGT-----CTTTTATTTATTTGATCTGTATTTGATTTTAAAGGCTT 1450
QY 1715 gtaataatgaatgtcaactagttgcaatataagaacaaattagctattttaaattatt 1774
Db 1449 ATATATCATGAAATATCATCTAGTGGCCAGAAATATTAATTAACACTGCTTTAATATGAAT 1390
QY 1775 aaccctgtcgaactcaattcagggtgcacatatactggtgtgaagctgaagcaagcc 1834
Db 1389 AATATGTATAGCTTAGGACTTTACTTTAGTTTCATATACCTGCTGCTAGTGGGCAACAT 1330
QY 1835 acaataaggagagctctatatta-cttataagacatacctaagaagatgaagctgaataga 1893
Db 1329 GAAGTATAGACAGACTTCTGTTGATTTTATTTAGGCGCATTAAGGAATGAGCTGAACAG 1270
QY 1894 cccctctatacccttctgcttaataatgaatgaataatctcagctgtttaaactc 1953
Db 1269 CCGCTGATACCTTTCTCTTAATTAACATGATGATTAATTCAGGTAACATTAACACCT 1210
QY 1954 gttttgtacaccttcccaaaaattattgtcatcagcaatccctgacatgtagttc 2013
Db 1209 GTTGTGTTTCACCTTCTCTCTATAAA-----ATTGTACAGCTCTCTGTGACACTTAACCT 1156
QY 2014 caaactttagctctccacagagctgagcagcactgtatcatcactgagcctgagcaactcac 2073
Db 1155 CAACCTTTAGCATCTCTGTGAGAGCTGCATCCACTGTATTAATTTCCCTGCGCAACTGAG 1096
QY 2074 tggaggaaagatgcccagagagctgcacatgtccctctctgcttgcaggaagcagtgcc 2133
Db 1095 TGAGGGGAATGTGCCAGAGGAGCTGCAAGCACTCCCTCCCTGCTTCAGGGTCAAGAGTG 1036
QY 2134 cagcactt-----aggcagcatccaagcagagggctcagcgcccaagggcttggacggtta 2186
Db 1035 CCCAGGCTTTATATAGAGGAGCATCCAGAGCCCGAGAGCACTGCTCTTCCGCTGGTG 976
QY 2187 tctctccct-99ggcgtgttaattgtgtgataagccctgagcccaagagagcagcgcg 2244
Db 975 CCTTCTCTCTGAGGGGCTATCATGTGTAGATTAAGCCCTGAGTAGAGCAAGACAGAG 916
QY 2245 atccacagtcagtgttccatgacccctc-tcccttcccttcccaagcaactgagatatt 2303
Db 915 ATCCACTGTCTATGTGCTGTATCATCTCAAACTTCCCTCCAGCAGAGGAATATT 856
QY 2304 gcttgcagtcataccgcaagaana 2329
Db 855 GCGTGGCATGCAACCTGCAAAAGAA 830
```

```
RESULT 13
AA159901
ID AA159901 standard: cDNA: 929 BP.
XX
AC AA159901:
```

```
XX 22-OCT-2001 (first entry)
DT
XX
DE Human polynucleotide SEQ ID NO 3890.
XX
KW Human: nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000MO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
XX
PT P-PSDB: AAM40745.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PS such as central nervous system injuries -
XX
PS Claim 1: SEQ ID NO 3890; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localized neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 929 BP; 214 A; 224 C; 223 G; 268 T; 0 other;
```

```
Query Match 22.6%; Score 642.4; DB 22; Length 929;
Best Local Similarity 83.4%; Pred. No. 2.8e-139;
Matches 742; Conservative 0; Mismatches 146; Indels 2; Gaps 1;
```

```
QY 5 ggaggcgagggtctcgcgctcccgctgagctctcttccgcgagcaggttgcgcc 64
Db 42 ggcccgggggcttcggcgctgctcgcgcgctccctcttcgcacagtgaggatcac 101
QY 65 gtatgcctctgaagcatatgaagtcctccggttccgacttagcccgagcagcgaga 124
```


| | | | |
|----|-----|---|------|
| Db | 484 | ggggaaatgtcgtctagatttggatcccttcggacgtcgtctcaacctgltcaatcc | 5433 |
| Qy | 545 | ccctcgtcgaattccgagtcggagcccttgltgcacccagggaactaaagggtcagg | 6040 |
| Db | 544 | catgtctcgagatttagagtgtagccacatcatgttaactcagagcactagaagactagg | 6030 |
| Qy | 605 | agaaggtgtcacatatccagccatgatgcacatgvtgtgtctcatatgggtccccccttga | 6644 |
| Db | 604 | agaaggtgttaccattccagcgaatgcacatgcatgtgtgtctctctgtgctccccccttga | 6630 |
| Qy | 665 | aagaagcgaagcttcgaattatctatattatgtgcaggagcacaactgtggcagtagttctctc | 7240 |
| Db | 664 | aagaagcgaacctctccagcatctcatatgcaggagcagaagctgtggcagtagtaattctct | 7230 |
| Qy | 725 | tcctcttctcgtagtaattgctactatatgaattgacatgactatgctctcatctcttgg | 7844 |
| Db | 724 | tcctcttctcgtgaataatttgctactatatgaattgacatgactatgctctctcaactttttgg | 7830 |
| Qy | 785 | catgtgtggaatcatctggtttattattatgagatcgtcttagttgatatcacccaagaac | 8444 |
| Db | 784 | nactattgaaaaaatttggtctctcttngngatctngtag-tagtgaacacacacaacaaa | 8422 |
| Qy | 845 | tcacaagaacaa | 855 |
| Db | 843 | accccaagaaa | 853 |

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
xx
SQ Sequence 838 BP; 189 A; 196 C; 210 G; 237 T; 6 other:

| | | | | |
|-----------------------|-----------------|---------------------|-----------|-------------|
| Query Match | 19.4%; | Score 553; | DB 22; | Length 838; |
| Best Local Similarity | 81.9%; | Pred. No. 1.5e-118; | | |
| Matches 660; | Conservative 0; | Mismatches 143; | Indels 3; | Gaps 2 |

| | | | |
|----|-----|---|-----|
| OY | 5 | ggggggggggggttcctggcggttcctcgctctgagctctctcttctccgggagcaagtttgcc | 64 |
| Db | 6 | ggccccggggggtcttcggcggttcggcgcgccctctctctctgccaagttgcagttacac | 65 |
| OY | 65 | gtagctccctgaaagcatcatalgaagttcccggtttctcgactttagcccgagcgagcgca | 124 |
| Db | 66 | ct -gtcaacgtaggcgttcatalgaggtctccggtttccgtagaacctggcccggaacgtaggga | 122 |
| OY | 125 | ggaggggtctcgagcgcacacgcgtctctgcagcgcccgcgggcggaaccgcgtccagt | 18 |
| Db | 124 | ggagagcgcgcggcccgcaagcctctctctccggcgcccgccacggcgcgcaagcgcgtccagt | 18 |
| OY | 185 | atgcgcctctgctgttctcaaaccttgcacatttttgcctctttttgtttctcttcgtctcta | 244 |
| Db | 184 | gtgcgcctctgcgtcttcacaacttgcacattttgcctctttttgtttctcttcattgtgtta | 24 |
| OY | 245 | tctattcagcgggtgaactctgcagcgttcgactagttggacatggltggtatltcaaacacactgc | 30 |
| Db | 244 | tgcattttagcgttgagatctcgagtggtgtcggttagtgatagtgttaagtttcaatacacttt | 30 |
| OY | 305 | caaaataatagaacgcttctcagagttgcgagagcatctctcctccataaaagtctctca | 36 |
| Db | 304 | agaaataataagaacttcccaagcgtgtgcagagagatctctcctccataaaagtctctca | 36 |
| OY | 365 | caaccgaacgggttaaaagtgaccggttggagtgcaagaaactcaagatgagttcttcogata | 424 |
| Db | 364 | taatcaaacgcggtaagaggtaccaatgtgagtgacgaagaaactcaagatgagttcttcogttc | 422 |
| OY | 425 | tttttctctatgctacatcaltcacacaacattctccttgagagatagtgttgcacagcaagaatgg | 484 |
| Db | 424 | ctttttttatgctacatcaltcacacaacagattctccggagatagtgttgcacagcaaaatagg | 485 |
| OY | 485 | gggggaagctgtgtgcagagatccggagatctttgtctcacagctatcttcaacctgttcacttc | 544 |
| Db | 484 | gggggaagatgctgcctagagatttggatctcttgcacgtcgtctccacacccgtttcaactcc | 543 |
| OY | 545 | ctctgcctcagaattcggagatctcgagccctgttgcacatccggcgacatgaaagggcttagg | 604 |
| Db | 544 | catgtctgcagatttagagagtttggaacctcatgttaactcagagacatgaaaggaactang | 603 |
| OY | 605 | agaaggtgtcacatatccagccatgacatgacatgtgtctctcaatgttggtcccccctbta | 66 |
| Db | 604 | agaaggtgttcatcattccagccatgacatgacatgtgtctctcttgggtcccccctbta | 66 |
| OY | 665 | aagaagcagaactctcgagattttcatatagcggagacaaacttggnaaagtaatttccct | 724 |
| Db | 664 | aagaagcagaactctttagcatttcatatagcggagacaaacttggnaaagtaatttctctc | 722 |
| OY | 725 | tctcttcttcggagtaatttgcataatataatgaatgtgactatgtctctatcttctgtg | 784 |
| Db | 724 | tincttcttcgaaataa-ttgctactataatgttggaactatgacttctactcttttgg | 785 |
| OY | 785 | catgtgtggaatcatcgtgttattt | 810 |
| Db | 783 | gactatttgaaatatttgggtctcttt | 808 |

Search completed: July 15, 2002, 13:53:18
Job time: 14938 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 11:45:11 ; Search time 4323.72 Seconds
(without alignments)
8877.851 Million cell updates/sec

Title: US-09-776-865-3

Perfect score: 2844
Sequence: 1 cccggggggggggggctctg.....ataaaaaaaaaaaaaaaaaa 2844

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: qb_est1:*
10: qb_est2:*
11: qb_hlc:*
12: qb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 725.2 | 25.5 | 985 | 9 | AL550137 |
| 2 | 508.6 | 17.9 | 734 | 10 | BI907284 |
| 3 | 505.2 | 17.8 | 626 | 10 | BF676817 |
| 4 | 504.4 | 17.7 | 801 | 10 | BI661062 |
| 5 | 476 | 16.7 | 559 | 10 | BM106101 |
| 6 | 469.6 | 16.5 | 787 | 10 | BG198416 |
| 7 | 468.6 | 16.5 | 700 | 10 | BE869819 |
| 8 | 460.2 | 16.2 | 736 | 10 | BI860521 |
| 9 | 451.6 | 15.9 | 518 | 9 | AM655673 |
| 10 | 448.4 | 15.8 | 711 | 10 | BI697765 |
| 11 | 433.6 | 15.2 | 729 | 10 | BE971208 |
| 12 | 426.6 | 15.0 | 477 | 10 | BM106103 |
| 13 | 416 | 14.6 | 904 | 10 | BF124137 |
| 14 | 413.8 | 14.5 | 676 | 10 | BE867611 |
| 15 | 409.2 | 14.4 | 633 | 10 | BI817031 |
| 16 | 407.4 | 14.3 | 895 | 10 | BF539146 |
| 17 | 399 | 14.0 | 459 | 10 | BF652406 |

| | | | | | | |
|----|-------|------|------|----|----------|----------|
| 18 | 395.8 | 13.9 | 682 | 9 | BB613552 | BB613552 |
| 19 | 380.6 | 13.4 | 518 | 9 | AA833297 | AA833297 |
| 20 | 375.8 | 13.2 | 689 | 10 | BG400588 | BG400588 |
| 21 | 374.6 | 13.2 | 434 | 10 | BF652410 | BF652410 |
| 22 | 368.8 | 13.0 | 645 | 9 | BB537525 | BB537525 |
| 23 | 365.4 | 12.8 | 432 | 9 | AA258513 | AA258513 |
| 24 | 359 | 12.6 | 630 | 9 | BB610013 | BB610013 |
| 25 | 346 | 12.2 | 629 | 9 | BB664731 | BB664731 |
| 26 | 335.2 | 11.8 | 965 | 10 | BG290613 | BG290613 |
| 27 | 324 | 11.4 | 661 | 10 | BG400668 | BG400668 |
| 28 | 322.8 | 11.4 | 1100 | 10 | BG866603 | BG866603 |
| 29 | 317.6 | 11.2 | 360 | 10 | BF042802 | BF042802 |
| 30 | 310.6 | 10.9 | 499 | 9 | AL597124 | AL597124 |
| 31 | 304 | 10.7 | 320 | 10 | BG687881 | BG687881 |
| 32 | 302 | 10.6 | 429 | 10 | BF563945 | BF563945 |
| 33 | 300.4 | 10.6 | 658 | 10 | BI851890 | BI851890 |
| 34 | 273.6 | 9.6 | 587 | 10 | BM179717 | BM179717 |
| 35 | 272.4 | 9.6 | 721 | 10 | BJ060500 | BJ060500 |
| 36 | 270.2 | 9.5 | 524 | 10 | BM272092 | BM272092 |
| 37 | 261.4 | 9.2 | 367 | 10 | H63685 | H63685 |
| 38 | 261 | 9.2 | 301 | 10 | N31254 | N31254 |
| 39 | 255.8 | 9.0 | 893 | 10 | BG541099 | BG541099 |
| 40 | 254.6 | 9.0 | 632 | 10 | BJ031727 | BJ031727 |
| 41 | 246 | 8.6 | 360 | 10 | BE663434 | BE663434 |
| 42 | 237.6 | 8.4 | 445 | 10 | BE375421 | BE375421 |
| 43 | 233.8 | 8.2 | 376 | 9 | AM486714 | AM486714 |
| 44 | 224.4 | 7.9 | 259 | 10 | BI536090 | BI536090 |
| 45 | 219.6 | 7.7 | 321 | 9 | AM531337 | AM531337 |

ALIGNMENTS

RESULT 1
AL550137 985 bp mRNA linear EST 16-FEB-2001
DEFINITION AL550137 LTI_NFL006.PL2 Homo sapiens cDNA clone CS01040YK17 5
LOCUS prime, mRNA sequence.
ACCESSION AL550137
VERSION AL550137.1 GI:12886813
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 985)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source

1. 985

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS01040YK17"

/clone_lib="LTI_NFL006.PL2"

/issue_type="Placenta"

/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA

was primed with a NotI-oligo(dt) primer. Five prime end

enriched, double-stranded cDNA was digested with NotI and

cloned into the NotI and Eco RV sites of the pCMVSPORT 6

vector. Library was normalized. Library was constructed by

Life Technologies. Contact : Feng Liang Life Technologies,

a division of Invitrogen 9800 Medical Center Drive

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifestech.com URL :

http://fulllength.invitrogen.com"

BASE COUNT 236 a 224 c 224 g 300 t 1 others

ORIGIN

| | | | |
|----|-----|---|-----|
| OY | 366 | aacccaagcggrraaaaagttaccgcggyggatgtaagaactcaagatgtatcttcggactc | 425 |
| Db | 359 | AATCAACGGGGTAACAAAGTACCAATGGGGATGACGAACCTCAAGGATGATTTCTGGTTCC | 418 |
| OY | 426 | ttttctatgctacatcatcacacaacttccttgtaagatagtgttgcacagaaagtgg | 485 |
| Db | 419 | TTTTTTTATGGCTACATCATCACAGATTCCTGGAGGATAGTGTGCCAGCAAAATAGG | 478 |
| OY | 486 | gggaagctgtgtctagatctcgggactccttgcctacagctatctcaacctgttcactcc | 545 |
| Db | 479 | GGGAAAGTCTGCTAGGCAATTGGGATCCTTGCGACATCGCTGCTCCACCTGTTCACTCC | 538 |
| OY | 546 | ctcgttgtagatttcgggagtcggagcccttgttcaactcagggacactagaagggctaga | 605 |
| Db | 539 | ATTGCTGCGAGATTTAGGAGTTGGACCTCATTTGACTAGGACACTGAAAGCACTAGGA | 598 |
| OY | 606 | gagggctgcacataccacgacatcatgcattgcattgtcttcaatggctccccccttga | 665 |
| Db | 599 | GAGGGTGTATCAATTCCAGCCCATGCAATGCAATGCTGTCTTCTGGGCTCCCTCTTGAA | 658 |
| OY | 666 | agaagcaagctcttgagiatatttcatalatgcagagagacaaacttggagactt-aggttctct | 724 |
| Db | 659 | AGAAACAAACTCTTAGCATTTTCATATGACAGAGACACAGCTTGGACAGTAAATTTCCT | 718 |
| OY | 725 | tcctctcttcctggagtaatttgcatactaatgaattgg | 761 |
| Db | 719 | TCCTCTTCTTGGAATTA-TTGGCTACTAATATGAATGG | 754 |

| | |
|------------|--|
| RESULT | 3 |
| B6F76817 | |
| LOCUS | B6F76817 626 bp mRNA linear EST 21-DEC-2000 |
| DEFINITION | G020P8438001 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4248811 5', mRNA sequence. |
| ACCESSION | B6F76817 |
| VERSION | B6F76817.1 GI:11950712 |
| KEYWORDS | EST. |
| SOURCE | human. |
| ORGANISM | Homo sapiens |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 626) NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) |
| AUTHORS | Unpublished (1999) |
| TITLE | Contact: Robert Strausberg, Ph.D. |
| JOURNAL | Email: cqapbs-r@mail.nih.gov |
| COMMENT | Tissue Procurement: CLONETECH Laboratories, Inc. cDNA Library Preparation: CLONETECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LICM1067 row: 0 column: 20 High quality sequence stop: 398. |
| FEATURES | Location/Qualifiers |
| source | 1..626 |

| FEATURES | SOURCE |
|---------------------|---|
| Location/Qualifiers | 1. 626 |
| | /organism="Homo sapiens" |
| | /db_xref="taxon:9606" |
| | /clone="IMAGE:4248811" |
| | /clone_lib="NIH MGC_83" |
| | /lab_host="DH10B (T1 phage-resistant)" |
| | /note="Organ: prostate; Vector: pNNR-LIB (Clontech); |
| | Site_1: STII (ggcgcgcctggcgc); Site_2: STII (ggccatataggc |
| |); 5' and 3' adaptors were used in cloning as follows: 5' |
| | adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor |
| | sequence: 5'-ATTCGAGAGCGCAGCGCGCCGACATG-dt(30)BN-3' |
| | (where B = A, C, or G and N = A, C, G, or T). Average |
| | insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies |
| | contained inserts by PCR. This library was enriched for |
| | full-length clones and was constructed by Clontech |
| | Laboratories (Palo Alto, CA)." |

| | | | | |
|------------|-------|-------|-------|-------|
| BASE COUNT | 157 a | 126 c | 123 g | 220 t |
| ORIGIN | | | | |

| | | | | |
|---------------------------|-------|------------------|----------|------------|
| Query Match | 17.8% | Score 505.2 | DB 10 | Length 626 |
| Best Local Similarity | 89.0% | Pred. No. 2.1e82 | | |
| Matches 557; Conservative | 0 | Mismatches 68 | Indels 1 | Gaps 1 |

| | | | |
|----|------|---|------|
| QY | 1001 | tactgtgtgacctattgctgaccttaactaaatgaagaagaaagtcctaaagttcaattatcaata | 1060 |
| DB | 1 | ttacttatttgacatttattggccttacttattatgtgaagagatccttaaggcttaattgttcaaa | 60 |
| QY | 1061 | gaatgggttttcatctgcgaagtcacctattataggttgtgtgtatgtatgatctctcg | 1120 |
| DB | 61 | gaatgggtttttattcttatttgccttattttaggccttggttatgtatgatgcctgcg | 120 |
| QY | 1121 | tcaagctgctgcaaatltaagggaagatgtaatlcttcaactctgtyggttcgaagct | 1180 |
| DB | 121 | tcaacctgctgcaaatTTAAAGGCAAAATGGAATTTTCAACTTATATGTTGTCGCAGAT | 180 |
| QY | 1181 | ttttgaccttaaggatgatgtgagccgcgcgatcttcctggtgttcgcgagattatag | 1240 |
| DB | 181 | TTTTAGCCTTATAGAGAAATGATTTGGACTGCTGATTTCTGGTAGCTGCTTGCTTATTGG | 240 |
| QY | 1241 | ctgtgattatctctgtgctgtgtgcaatctctaaccatacaaacccctggaagctttg | 1300 |
| DB | 241 | CTGTGATTATTCTTGGGCGTTGCTTCTTCAATATCAACAAACATGGGAGCGTTTGG | 300 |
| QY | 1301 | ctctcttgatttgatcaatcaacatctggaatctgctccttcgtatgctgtatctcct | 1360 |
| DB | 301 | CTCTTCTGGATTTTACCATTCACCACTCTGGATATWCTCTCTGATATGCTGTATCCTCT | 360 |
| QY | 1361 | gggcaatcaaaatccttttgcacatctccgtgaatgtatgtggcccatgtgcagaga | 1420 |
| DB | 361 | GGGCATTCACAAATACATTTCGCATATTCAGAGATGGTGGCCCGCATTTGCTAAAG | 420 |
| QY | 1421 | tcttaaccctcgaaacactatgtggaatgtgcaaatgttctgtcatcgtcgtcat | 1480 |
| DB | 421 | TCTGACCCCTGTAAACAGTGTGGAGAAATGGCAAAACCGTCTCTATATTTGGCTGCTAT | 480 |
| QY | 1481 | caatgatcttggtgcatttcttcaacatctgcgaagttgaagtcgaactgggc | 1540 |
| DB | 481 | TATATGTTTTGTGTCATTTCTTTACACTATTCGCCAAAGGTGAATGACAAATGTGGC | 540 |
| QY | 1541 | catcgtgatcccaaggacacagaagaactgaagaaacaat-aaataatcctgtctat | 1599 |
| DB | 541 | TCTCAATGATACCAATGACACAGACACTGAAGAACAAATTAATATTCCTGCTGTAT | 600 |
| QY | 1600 | taatgatccttggttatcaatgaac | 1625 |
| DB | 601 | TTTACGTATTATTATATCATGTGTAAC | 626 |

| | |
|------------|---|
| RESULT | 4 |
| LOCUS | B1661062 |
| DEFINITION | B1661062 801 bp mRNA linear EST 12-SEP-2001 603304362n1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:535004 5', mRNA sequence. |
| ACCESSION | B1661062 |
| VERSION | B1661062.1 GI:15757298 |
| KEYWORDS | EST. |
| SOURCE | house mouse. Mus musculus |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. |
| REFERENCE | NIH-MGC http://mgc.nci.nih.gov/ . 1 (bases 1 to 801) |
| AUTHORS | National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) |
| TITLE | Contact: Robert Strausberg, Ph.D. |
| JOURNAL | Email: cgapbs@mail.nih.gov |
| COMMENT | Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth Ph.D. |

Db 1 TAAATAATTCACAGGCTTCTGTTAAACATCTGTTTTGTACACCTTCCTCAAAAAA-TATTT 59
Qy 1985 gtcatacgaacatccctgacatgtagtctcaaaccttagcctccacaggaactggcagc 2044
Db 60 GTCATCAGAAATCCCTGACACATAGTGTCAAACTTTAGCCTCTTGAGCTAGCTGGCAGC 119
Qy 2045 cactgtatcatcagcctgacactcactcaggaagaagatgccagagcagctgcacat 2104
Db 120 CACTGTATCATTCAGCCCGGCAACTTCACTGAGGAGAGATGCCAGGAGAGCTGCCACAC 179
Qy 2105 gtccctctctgtgtcctcaggaagcagtgccagcacttagcagcatccaagaagagtc 2164
Db 180 GTCCCTCTCTGTGGCTTCAAGGAGAGTGGCCAGCACTTAGGCGCATTCACAAGACCGGGTC 239
Qy 2165 agcgcgaagagcttgagcagtgatcttccctcgggagctgtatgtgtgagtagagcct 2224
Db 240 AGTGGCAAGGCTTTGGATGTATTTCTTCCCTGGGGCTGTATATGTATGATGAAGCCCT 299
Qy 2225 gaggcaacaggaacagcgacgtacacagtcacatggttccatgacccctccctccctc 2284
Db 300 GAGCCAAAGGAGACAATCTGATCCACAGTCATGATTCATCCATCCCTCCCTCCCT 359
Qy 2285 cccagacacactggaatgtgctgtgacatgtaacctgcaaaagaagtgtgacttaatt 2344
Db 360 CCCAGC-----GCATGTATCTGCAAAAAGAAAGTGTGATGCTTAAT 401
Qy 2345 agccacataacatcatcctctgtagtaccctcactcactcactcagatgataatcctt 2404
Db 402 AGCCACATATACATCATCTTGTATGATCTTACCTTCACATGATGATCAAGATTAATCTT 461
Qy 2405 caagtcctgtgtctcaggaagctacacacgaataataataaagaagaacaaaca- 2463
Db 462 CAAGCTGTGTGTGTAGAGAGTACACACGAATATTAATAATTAATAAGAAACAT 521
Qy 2464 ttttctgtctgacactaagtgctgtgtgagctc 2501
Db 522 TTTTGTGTGTGTACACCTTAAGTATCTGTGTGCACTTC 559

RESULT 6
Bg198416 787 bp mRNA 1linear EST 21-APR-2001
LOCUS Bg198416
DEFINITION RST17674 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION Bg198416
VERSION Bg198416.1 GI:13720103
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 787)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thorndon,M., Ramachandran,R., Whittington,J.,
'Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith,
'E., Veloso,N., Kluka,A., Hess,J., Cothen,K., Lo,K., Offenbacher,
'J., Danzig,J. and Ducar,M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
TITLE JOURNAL
MEDLINE 21227151
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scaine@atersys.com
High quality sequence stop: 549.
Location/Qualifiers
1..787
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"

FEATURES
Source

/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 214 a 166 c 154 g 253 t
ORIGIN

Query Match 16.5%, Score 469.6; DB 10; Length 787;
Best Local Similarity 78.5%; Pred. No. 6.5e-76;
Matches 620; Conservative 0; Mismatches 149; Indels 21; Gaps 4;

Qy 1343 gtagctgtagtattccctcgggagatcacaaatccttgcacattccttgatgattgg 1402
Db 3 GATCTGTGTATCTCTCGGGGATCACAAATACATTTCCACTATTTCCAGGAATGGTTGG 62

Qy 1403 gccatcatgccaagaagcttaccctcctgagagacactatggaagatgcaaacgtttt 1462
Db 63 GCCCGTCAATGTCTAAAGCTGACCCCGATACACTGTGGAGATGGCAACCGTGT 122

Qy 1463 ctgcatcgtcgtctgctatcaatgatttgggccaatttcttcacactatccgaag 1522
Db 123 CCATATGTCTGTGCTATTAATGTTTTGGCCCATTTTCTTTACACATATTCGCCAAGG 182

Qy 1523 tgaagtgcgaactggccatcagtgatcacagaagcaacgaactgaagaacata 1582
Db 183 TGAAGTACAAACCTGGGCTCTCAATGATCACCATGACACGACCTGAAGAACCAATA 242

Qy 1583 aataatcctgtctcattaaatgatacttgttatacatgtaaacctaaagtgccttggat 1642
Db 243 AATATTCCTGCGCTCATTAATGATATTTTATTAATCAATGTAACCTCAAGTGCCTT----- 298

Qy 1643 atttaatgtgtaagaactatataacagaabaaattgtactagaataattgtgtaga 1702
Db 299 -CTGTATGTGTGAACCATATGATGTATGT-----CTTTTATTAATGTGATGTATGAGA 348

Qy 1703 ttgtgaagctgtgataatgaatgaatgactagtgccatlaagcaaatgattt 1762
Db 349 TTTTAAAGGCTATATATATGATGAATATATCATGATTTGCCAGAAATTAATAAGAACTGTGT 408

Qy 1763 ttaattatataacccgttctgtagaactacaactcaggaaggtcacatatcgtgtgcaa 1822
Db 409 TTAATTTAATGAATATATGTAAGCTAGACTCTACTTTAGGTGCATACATACCTGCTGCTA 468

Qy 1823 gtcaaggcaaccocaataagggagttctattta-ttataagacacatacctaagaagt 1881
Db 469 GTGGGGCAACATGAAGTGAAGAGAGTTCGTGTGATTTTAAAGGCCATGCTAAAGGAATG 528

Qy 1882 agctgaatagagccctctataccttgccttaattgaatgaatgaatgaatgaatgaat 1941
Db 529 AGCTGAACACAGACCTCCCTGATACCTTTGCTTAATTAACCTGATGATTAATTCAGGATC 588

Qy 1942 tgttaaacatctgttcttgaacaccttcccaaaaaatattatgtfcaacgaacacctgt 2001
Db 589 TGATTAACACCTGTGTGTGTACACTTGTGCATATAA-----ATTGACACTGCTGCTG 642

Qy 2002 acatgtaggtccaacttagcctctccacggagcgtggcagcaactgataatcagcc 2061
Db 643 ACGCTTAGACCTCAAAACTTTAGCATCTCTGTGGAGCTGCCATTCATATTAATTTTGGCC 702

Qy 2062 tggcaactcctcagggaggaacatgcccagcagctgacacatgtccctctcgtgctc 2121
Db 703 TGGCAACTGGACTGAGGGAGATGTGCCAGGACGCTGCCAAGCACTCCCTTCGTGGCTTC 762

Qy 2122 agggacagtg 2131
Db 763 AGGGTCAGAG 772

RESULT 7
BE869819

LOCUS BE869819 700 bp mRNA linear EST 20-OCT-2000
DEFINITION 60144665JF1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3850740 5',
mRNA sequence.
ACCESSION BE869819
VERSION BE869819.1 GI:10318595
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 700)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ARCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM9570 row: e column: 13
High quality sequence stop: 697.
Location/Qualifiers
1..700
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3850740"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 185 a 141 c 130 g 244 t
ORIGIN

Query Match 16.5%; Score 468.6; DB 10; Length 700;
Best Local Similarity 84.2%; Pred. 1e-75;
Matches 592; Conservative 0; Mismatches 94; Indels 17; Gaps 5;

QY 1106 taatactctcgaggaagctttagacctgacatggaaggaaggaatttcaactc 1165
|||||
Db 1 TATGATCTGCTGCTGAACCTGCTGACAAATTAAAGGCAAAATTTTCAACTTT 60
QY 1166 gtgagtcgaagaagctttagacctatagaaggaatggaacctgacatctgtgc 1225
|||||
Db 61 ATGATGTCGCGAAGATTTTATGACCTATAGGAATCATTTGACCTGACATTTCTCGTAGC 120
QY 1226 cgcagaattatagagctgtgattatctctgtgctgtgcatctcctaacaacaac 1285
|||||
Db 121 TCGTCGCTCATGCGCTGATATTTCTTGGCGCTTCTTCTTACATATCAACAC 180
QY 1286 ccgagagagctttagctctctgattatgacatcaacacatggaacctgtcctctgta 1345
|||||
Db 181 ACTGGGAGGCTTTGCTCTCTGATTTACATCAACCACTGGATTTCTCTCTCTGTA 240
QY 1346 tgcgtgattctctgaggaatcaacaatcttgcacacatctcctggaatgtaggc 1405
|||||
Db 241 TGTCTGATCTCTCTGCGCATCAACATATTTGCTTCCATTTCCAGGAATGTTGGGCC 300
QY 1406 catcatgtcagaagatcttaacctgagaacacatattgagaatggaactgtttctg 1465
|||||
Db 301 CGTCATTTGCTAAAGTCTGACCCCTGATACAGTGTGGAGATGGCAACCGTGTCTA 360
QY 1466 catcgct-gctgtctataatgattgtgtccatttcttcaacactattgcgaaggtg 1524
|||||
Db 361 TATTGCTCGCTGCTATTAAATGTTTGTGTCATTCTTACACATTTATGCGCAAAAGTG 420

QY 1525 aagtcgaagaactgagccatcagtcacaccagaagacagaagaactgaaggaacaaataa 1584
|||||
Db 421 AAGTACAAACTGGGCTCTCAATGATCACCATGACACAGACACTGAGGAACCAATPAA 480
QY 1585 taatcctgtctataatgattatctgtttatcatcagtaaccataagtgcccttgatat 1644
|||||
Db 481 TAACTCGCTCTATTAAATGATTATTTATTTATCATGTAACCTCAAAAGTCCCTT-----C 535
QY 1645 tttaattgaagaacatctatatacaagaataaaattgtaaaaaattgttagatt 1704
|||||
Db 536 TGTATTGTGTAAGCATTTCTATGT-----CTTTTATTATTTGTAATTGATTT 586
QY 1705 tgaaggtctgtataatgaagaatgcaact-agtgcacataaagaacaaattagctatt 1763
|||||
Db 587 T-TAAGGCTTATATCATGTAATATATCATCTGATGCCAGAAATATPAAATGAACGTGTT 645
QY 1764 taattattataaccgcttgcctggaacttaacattcaaggtc 1806
|||||
Db 646 TACTTATGAATPAAATATGTAAGTACGACTTCTTAACCTTAGGCTC 688

RESULT 8

B1860521
LOCUS B1860521 736 bp mRNA linear EST 10-OCT-2001
DEFINITION 60338678JF1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5395772 5',
mRNA sequence.
ACCESSION B1860521
VERSION B1860521.1 GI:16001268
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 736)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LHAM12008 row: m column: 21
High quality sequence stop: 713.
Location/Qualifiers
1..736
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5395772"
/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 216 a 142 c 147 g 231 t
ORIGIN

Query Match 16.2%; Score 460.2; DB 10; Length 736;
Best Local Similarity 87.4%; Pred. No. 3.4e-74;
Matches 515; Conservative 0; Mismatches 73; Indels 1; Gaps 1;

QY 318 agctctacgagtgctgacgacatctgtctccataaagttctcaacaacaacaggt 377
|||||
Db 1 ACTTCCAAAGCGCTGCTCAGAGCATTTCTGCTCCATTAAGATTCATCATATCAAAAGGGT 60

| FEATURES | Source | Location/Qualifiers |
|---------------------------|--|---|
| | | 1..518 |
| | | /organism="Bos taurus" |
| | | /db_xref="taxon:9913" |
| | | /clone_11b="MARC 1B0V" |
| | | /tissue_type="pooled" |
| | | /lab_host="DH10B" |
| | | /note="Vector: pCMV SPOR6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary." |
| BASE COUNT | 136 a 139 c 103 g 139 t | 1 others |
| ORIGIN | | |
| Query Match | 15.9%; Score 451.6; DB 9; Length 518; | |
| Best Local Similarity | 92.3%; Pred. No. 1.4e-72; | |
| Matches 492; Conservative | 0; Mismatches 25; Indels 16; Gaps 1; | |
| Qy | 1876 gagatgagctgaataagacccttcctatacccttctgttaattaagtgataataattctc | 1935 |
| Db | 2 GAGAGGACCTGAAAAGAGCCCTCTATACCTTTGCTTAATTAAGGTGATATAATTTCTC | 61 |
| Qy | 1936 aggtcttgtaaacaatcgtttttgttaacaccttcctcaaaaattattgtcatcaagaa | 1995 |
| Db | 62 AGGCTCTTTAAACATCTGTTTTTGTAAACCTCTCTCAAAAATTAATTTGTTCATAGCAA | 121 |
| Qy | 1996 tccctgacatgtagagttccaaactttaagccttcacagagagcttgagccacatgtatcat | 2055 |
| Db | 122 TCCCTGACACATAGGTCTCAAACTTTACCTCTTGACGTAGCTGGCACCACTGTATCAT | 181 |
| Qy | 2056 tcaagcttggaacttactactgagggaagacatgcccagagcagctgcacatgtccctctc | 2115 |
| Db | 182 TCAGCCCCGCAACTTCACCTGAGGGAGAGCATGCCAGCAGCATGCCACACAGTCCCTCTCT | 241 |
| Qy | 2116 ggccttcaggagcagagcccaagcacttaagcagcatccaagaccagagtgccagcccaagc | 2175 |
| Db | 242 GGCCTTCAGGAGCAGAGCCACACACTTAGGCGAGCATCCAAAGACCAAGGTCAGTGGCCAAAGC | 301 |
| Qy | 2176 tttagcaggtattcttcctccctggggctgttaatgtgtgtatgaagcccttgagccaacag | 2235 |
| Db | 302 TTTGGATGATATTCCTCCCTGGGGCTTTAATGTATGATGAATGAACCTCGAACCAACAG | 361 |
| Qy | 2236 gacagcgcgatccacagtcagtgttccatgcagcccttccttccttcccaagacact | 2295 |
| Db | 362 GACAACTGATGCCACAGATCATGATTTCCATGCACCCCTCCCTCCCTCCATCCAGCGCATG | 421 |
| Qy | 2296 gggagatgtgctgcatgttaacctgcgaaaagaagtgtgatggccttaattagccaatat | 2355 |
| Db | 422 TTAATCTT-----GCAAAANGAAATGTGATGCCCTTAATTAAGCCACATATTA | 465 |
| Qy | 2356 aacatcatccttgatgatcctaccctcacatgcatgataagatataactcttcaag | 2408 |
| Db | 466 ACATCATCTCTTGATGATCTCTACCTTCACATGATGATGATGATATAAATCTTCAAG | 518 |
| RESULT 10 | | |
| LOCUS | B1697765 | 711 bp mRNA linear EST 18-SEP-2001 |
| DEFINITION | 60334685F1 NCL_CGAP_Mam2 Mus musculus CDNA clone IMAGE:5374407 5' | |
| ACCESSION | B1697765 | mRNA sequence. |
| VERSION | B1697765.1 | GI:15660394 |
| KEYWORDS | EST. | |
| SOURCE | house mouse. | |
| ORGANISM | Mus musculus. | |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| AUTHORS | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | |
| TITLE | 1 (bases 1 to 711) | |
| JOURNAL | NIH-MGC http://mgc.nci.nih.gov/. | |
| COMMENT | National Institutes of Health, Mammalian Gene Collection (MGC) | |
| | Unpublished (1999) | |
| | Contact: Robert Strausberg, Ph.D. | |
| | Email: cgapds-tr@mail.nih.gov | |

Db 241 TTTTATTTATTCATGTAACCTCAAGTGCCTT-----CTGTATTGTGTAAAGCATTTCTANG 295
 QY 1667 tacaagataaaattgtactagaanaaattgtgttagatttctgaagctctgtacatcatgaa 1726
 Db 296 T-----CTTTTATTTATTTACTTGTATAGATTTTAAAGGCTATATCATGAA 346
 QY 1727 tgcctactagttgcacataaagaanaattagctatctttaaattatattaaaccggttgc 1786
 Db 347 TATACATAGTTGCCAGAAATATAAATGAACCTGTATTATATGATATATATGTAAGCT 406
 QY 1787 ggaactacatcaggtgcacatctgtcgaagctcgaagtaagcaaccacaataaggagag 1846
 Db 407 AGGACTCTTACTTGTAGGTTACATACCTGCTGTAGTGGGCAATGAAGAGAGAG 466
 QY 1847 ttcatatta-ttataagac 1905
 Db 467 TTTGTGTTGATTTTGTAGGCGCATCTAAAGGATAGAGTGAAGACAGACTCCTGATAC 526
 QY 1906 ttgtcttaattaaagtggataaattctcaagctctgtttaaaccatctgtttgtacac 1965
 Db 527 TTTGCTTAATTAACATAGATGATATTTCTCAGGTACTGATTAACACCTGTTGTTCAC 586
 QY 1966 ctccccaanaaattatttctcagcaatccctcagacatgtagtctcaactttagcc 2025
 Db 587 TTTCTCATTAATAA-----ATGTCTCAGCTCTCTCTGACACTTAGACCTCAACCTTAGCA 640
 QY 2026 tctccacagagctgcagcagcactgtatcattcagcctgcagcactcagcaggaagcat 2085
 Db 641 TCTCTGGAGGCTGCCATCCACATGTAATTTTGCCCTGCGCACTGACATGAGGAGACTGT 700
 QY 2086 gcccaaggcag 2095
 Db 701 GCCCAGGCGAG 710
 RESULT 12
 LOCUS BM106103 477 bp mRNA linear EST 21-NOV-2001
 DEFINITION 509742 MARC 3BOV Bos taurus CDNA 5', mRNA sequence.
 ACCESSION BM106103
 VERSION BM106103.1 GI:17037173
 KEYWORDS EST.
 SOURCE cow.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 REFERENCE 1 (bases 1 to 477)
 AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
 G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
 Perlea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
 Keefe,J.W.
 TITLE Sequence evaluation of four pooled-tissue normalized bovine CDNA
 libraries and construction of a gene index for cattle
 JOURNAL Genome Res. 11 (4), 626-630 (2001)
 MEDLINE 21180013
 COMMENT Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -mismatch 18
 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACACAGCTATGACACAT
 BACKWARD: GTTTCACAGTCAGCAGC
 Plate: 102 row: K column: 23
 Seq primer: ATTAGGTGACACTATAG.
 FEATURES
 source Location/Qualifiers
 1..477

/organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 3BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from marrow, alveolar
 macrophage, ovary, fetal semitendinosus muscle, and fetal
 longissimus muscle."
 BASE COUNT 121 a 134 c 93 g 129 t
 ORIGIN
 Query Match 15.0%; Score 426.6; DB 10; Length 477;
 Best Local Similarity 93.5%; Pred. No. 5.1e-68;
 Matches 463; Conservative 0; Mismatches 14; Indels 18; Gaps 1;
 QY 1925 taataattcgaagctctgtttaaacaatcgttttctgaacaccttcctcaaaaattatt 1984
 Db 1 TAATTAATTCAGAGCTCTGTTAAACATCTGTTTGTACACCTTCCCAAAAATTAATTT 60
 QY 1985 gtcatagaatccctgcacatgtagtctcaaaacttagctctcaggaagctgcagc 2044
 Db 61 GTCATCAGCAATCCCTGACACATGAGCTCAAACTTTAGCTTGTACGTAGCTGGCAGC 120
 QY 2045 cactgtatcatcagcctgcagcactcactgaaggaagcagccagcagctgcacat 2104
 Db 121 CACTGTATCATTTAGGCCGCCGCACTTACAGAGGAAGCATGCCAGCGCTGCCACAC 180
 QY 2105 gtccctctctgtgcttcaaggaagcagctgcagcacttagcagcactcaagaccaggtc 2164
 Db 181 GTCCCTCTCTGTGCTTCCAGGAGCAGTGCCAGCAGCTTAGCAGCATCCAGACAGGCTC 240
 QY 2165 agcgcaaggcttgcagcagctattctcccttgagcctgttaatgtgtgaatgaagcct 2224
 Db 241 AGTCCAGAGCTTTGGATGATATCTCCCTGGGCTGTAATGTGATGATGAAGCCT 300
 QY 2225 gaagcaacaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2284
 Db 301 GAGCACAACAGGAGCAACTGATCCACAGTCATGATTTCCATGACCCCTCCCTCCCT 360
 QY 2285 ccagcagcactgtagtattgtcctgcagctgaacctgcagaaagaagtgtgactaat 2344
 Db 361 CCCAGC-----GCATGTAACTCTGCAAAAGAAAGTGTATGCTCTAAT 402
 QY 2345 aggcacatataacatcctctgatgatcctactcactcagcagcagcagcagcagcagc 2404
 Db 403 AGCCATATATACATCATCTTATGATGATCTTACCTTCATGATGATGATGATGAT 462
 QY 2405 caagctcgtgtctc 2419
 Db 463 CAAGTCTCTGTCT 477
 RESULT 13
 LOCUS BF124137 904 bp mRNA linear EST 24-OCT-2000
 DEFINITION 60176083BF1 NCL_CGAP_Mam5 Mus musculus CDNA clone IMAGE:4024064 5',
 mRNA sequence.
 ACCESSION BF124137
 VERSION BF124137.1 GI:10963177
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 904)
 AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 08:22:15 ; Search time 32.66 Seconds
(without alignments)
1576.971 Million cell updates/sec

Title: US-09-776-865-2

Perfect score: 2836

Sequence: 1 MAAGAMTPRPYQPARPGF.....LFAKGEVQWMLNDHHGHRH 536

Scoring table: BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 987 | 34.8 | 493 | 2 | protein C38C10.2 (|
| 2 | 967 | 34.1 | 472 | 2 | hypothetical prote |
| 3 | 936 | 33.0 | 560 | 2 | brain specific Na+ |
| 4 | 822.5 | 29.0 | 465 | 2 | sodium/phosphate t |
| 5 | 809.5 | 28.5 | 465 | 2 | sodium-phosphate t |
| 6 | 808 | 28.5 | 563 | 2 | probable sodium-de |
| 7 | 808 | 28.5 | 576 | 2 | protein 2K512.6 (l |
| 8 | 801.5 | 28.3 | 467 | 2 | sodium phosphate t |
| 9 | 797.5 | 28.1 | 573 | 2 | hypothetical prote |
| 10 | 786.5 | 27.7 | 465 | 2 | Na+-dependent phos |
| 11 | 767.5 | 27.1 | 544 | 2 | hypothetical prote |
| 12 | 747 | 26.3 | 512 | 2 | hypothetical prote |
| 13 | 629.5 | 22.2 | 501 | 2 | hypothetical prote |
| 14 | 562 | 19.8 | 413 | 2 | protein F2566.7 (i |
| 15 | 531 | 18.7 | 537 | 2 | hypothetical prote |
| 16 | 530.5 | 18.7 | 466 | 2 | hypothetical prote |
| 17 | 492.5 | 17.4 | 561 | 2 | hypothetical prote |
| 18 | 486.5 | 17.2 | 568 | 2 | hypothetical prote |
| 19 | 476.5 | 16.8 | 592 | 2 | CO2C2.4 protein - |
| 20 | 460.5 | 16.2 | 516 | 2 | hypothetical prote |
| 21 | 452 | 15.9 | 530 | 2 | hypothetical prote |
| 22 | 445 | 15.7 | 380 | 2 | hypothetical prote |
| 23 | 444.5 | 15.3 | 543 | 2 | hypothetical prote |
| 24 | 432.5 | 15.2 | 499 | 2 | hypothetical prote |
| 25 | 431 | 15.2 | 506 | 2 | hypothetical prote |
| 26 | 419.5 | 14.8 | 493 | 2 | hypothetical prote |
| 27 | 392.5 | 13.8 | 478 | 2 | hypothetical prote |
| 28 | 389.5 | 13.7 | 445 | 2 | hypothetical prote |
| 29 | 388.5 | 13.7 | 462 | 2 | hypothetical prote |

| | | | | | |
|----|-------|------|-----|---|--------------------|
| 30 | 386.5 | 13.6 | 473 | 2 | hypothetical prote |
| 31 | 380.5 | 13.4 | 425 | 2 | hypothetical prote |
| 32 | 372 | 13.1 | 420 | 2 | membrane transport |
| 33 | 370.5 | 13.1 | 428 | 2 | hypothetical prote |
| 34 | 362 | 12.8 | 485 | 2 | hypothetical prote |
| 35 | 360 | 12.7 | 455 | 1 | probable glucarat |
| 36 | 357.5 | 12.6 | 455 | 2 | hypothetical prote |
| 37 | 357 | 12.6 | 452 | 2 | hypothetical prote |
| 38 | 355.5 | 12.5 | 659 | 2 | hypothetical prote |
| 39 | 353 | 12.4 | 516 | 2 | hypothetical prote |
| 40 | 349 | 12.3 | 450 | 2 | probable glucarate |
| 41 | 349 | 12.3 | 450 | 2 | probable transport |
| 42 | 349 | 12.3 | 450 | 2 | probable transport |
| 43 | 339.5 | 12.0 | 445 | 1 | hypothetical prote |
| 44 | 336.5 | 11.9 | 493 | 2 | hypothetical prote |
| 45 | 334 | 11.8 | 487 | 2 | hypothetical prote |

ALIGNMENTS

| Query Match | Score | DB 2; | Length | 493; |
|-----------------------|-------|--|----------|------------------------------------|
| Best Local Similarity | 39.7% | Pred. No. | 8.2e-71; | |
| Matches | 194; | Conservative | 94; | Mismatches 179; Indels 22; Gaps 6; |
| QY | 55 | ESTDRTPLPGAPRAEAAPVCCSARYNLALAFPFYVYALRYNLVALVDVDSNTT | 114 | |
| DB | 2 | EGATTKRPLVP-----STRFALSLVMFEGCLVTYMMRMTMSFAVVCVNMENKT | 49 | |
| QY | 115 | LEDNNTSKACPEHSAPIVYHNQTKYQMDAETOGMLGSEFFYITITDPCGVASKI | 174 | |
| DB | 50 | DTGVEKVRSGCKEMPRVENSNSVIG-EPDWKQKTGMVLSSEFYIGISQIIGHLSRY | 108 | |
| QY | 175 | GKKMLGFGIIGTAVLTFTPLAADLGAGPLIVBALGSGVYFPPMHMMWSWAPPL | 234 | |
| DB | 109 | GKKRVFVYITLGSALLTLNPLNPAARTSEYALALIAAGTLOGAATFPMHTMWSWGPP | 168 | |
| QY | 235 | ERSKLISYVAGQVGTIVSLPLSGIICY--NMNTVYFEFFGTIGFWMLTWLWSSD | 291 | |
| DB | 169 | ELSVLTGVTYVAGQVGNIVPLPSGLCEYFGDGMPSIFIIIGVGLVMAVMMVYVSSD | 228 | |
| QY | 292 | TPQKKRISHYEKETIYLSLRNQSQ---KSPVWVILKSLPLMAIVAAHFSYNNTFY | 347 | |
| DB | 229 | KPATPRITPEKQYIVAVASMGKDKVSPWIKILTSPAVMAWACHFAGDMGAY | 288 | |
| QY | 348 | TLTLTLPTPKMEILRFNVQENGPLSLPLGSMICMLISGQADNLRAKWNFTICVARI | 407 | |
| DB | 289 | TMLVSLPFLKDVGLNLSSIGAVASTPIYAFYLAINGVGLADLRKSGILSTLNTARRA | 348 | |
| QY | 408 | FSLGIMGPAVFLVAAAGFIGDYS-LAAVFLITSTLLOGFCSSGFSINHLDIASVAGIL | 466 | |

Db 349 AMVALIGGIFLVASGCGGDDVLIITTCGMAISGLQVAGFVNVLEIAPPESGV 408
 QY 467 LGITNFATIPGAVPIAKSLTPDNTVGEMQVIFYIAAINVEGAFITFLPAKGEVOM 526
 Db 409 MGNNTISALAGIISPAVSSYLTPNGQEMQVWLMTAGILITIGALLFIFASGEVQPM 468
 QY 527 A-LNDHHGH 534
 Db 469 AKLAEEGH 477

RESULT 2
 S28286
 Hypothetical protein C38C10.2 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 16-Feb-1997
 C:Accession: S28286
 R:Thomas, K.
 Submitted to the EMBL Data Library, December 1992
 A:Reference number: S28285
 A:Accession: S28286
 A:Molecule type: DNA
 A:Residues: 1-472 <THO>
 A:Cross-references: EMBL:219153
 C:Genetics:
 A:Introns: 50/3; 287/3; 351/3; 412/3
 C:Keywords: transmembrane protein

Query Match 34.1%; Score 967; DB 2; Length 472;
 Best Local Similarity 41.0%; Pred. No. 3,1e-69;
 Matches 187; Conservative 87; Mismatches 172; Indels 10; Gaps 5;

QY 88 FFEFFIYALRVMLSVALVMDVSNFTLEDNRTSKACPEHSAPIKVHHNQTGRKKYQDAE 147
 Db 2 FFGCLVYIMARKTMSPAVVCAMVNEKTDIGVEKSRGKEMTPVESSSVIG-EPMDKQ 60
 QY 148 TQGMILGSFFGYIITQIPGGVYASKIGKMLGFGILGTAULTFTPIADLVGPLYV 207
 Db 61 TTGMVLSSFFGYIGSQIGIIGHLASRGKRVYFVITLGSALLTLNPAVARSEVALAI 120
 QY 208 LRALBEGIEVTPPAHAMAMSSWAPLERKSLISTYAGQDGTIVSLPSITCIY--- 264
 Db 121 LRAIIGLOGATPPAMTMSVMPLELSVLGVYAGQIGNVTVLPSIGLCEGFD 180
 QY 265 MNNTYVEFGFTGIFMFLMVLVSDTPQKHKRISHYEKEYLISLRNQLSSQ---KS 320
 Db 181 GGRPSIFITIGVGLMTAVMTVSSDKPATRPRITPEEKQYIVTAEASMGKDTGKPS 240
 QY 321 VPWVPIKSLPLMAIVVAHPSYMTFTYLLTLPTYMKELIRPNVOENGLSLPYLGSW 380
 Db 241 TPWIKILTPAVVACMGHAGDMGAVTMLVSLPSFLKDVLTGLNLSLGAVASIPYIAF 300
 QY 381 LCMILSQADNIRAKNFTSLCYRIRFSLIGMIGPAVFLVAAAGTIGDYS-LAVAFLT 439
 Db 301 LAINAGVGLADLRSGKISTLNTRRAMVALIGIGIFLVASGCGGDDVLIITTC 360
 QY 440 STILGCGSGFSINHLDIAPSYAGILGTTFATIPGAVGVIAKSLPNDTVGEMQV 499
 Db 361 GMAISGLQVAGFVNVLEIAPPESGVYMGNTISALAGIISPAVSSYLTPNGQEMQ 420
 QY 500 VFYIAAINVEGAFITFLPAKGEVQVMA-LNDHHGH 534
 Db 421 VMLTAGILITIGALLFIFASGEVQPMAKLAEEGH 456

RESULT 3
 I59302
 brain specific Na+-dependent inorganic phosphate cotransporter - rat
 C:Species: *Rattus norvegicus* (Norway rat)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
 C:Accession: I59302
 R:NI, B.; Rostek, P.R.; Nadi, N.S.; Paul, S.M.

Proc. Natl. Acad. Sci. U.S.A. 91, 5607-5611, 1994
 A:Title: Cloning and expression of a cDNA encoding a brain-specific Na(+)-dependent
 A:Reference number: I59302; MUID:94261635
 A:Accession: I59302
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: mRNA
 A:Residues: 1-560 <RES>
 A:Cross-references: EMBL:U07609; NID:g507414; PIDN:AAA1646.1; PID:g507415

Query Match 33.0%; Score 936; DB 2; Length 560;
 Best Local Similarity 39.0%; Pred. No. 1,1e-66;
 Matches 190; Conservative 91; Mismatches 168; Indels 38; Gaps 7;

QY 53 DGEESTDRTP---PLLP-----GAPRAEAPVCCSARNLALAFGFFIYALRVNLSVAL 105
 Db 38 DGRPVYTHTRDPVYVCTCGRLP-----RTIILMSGLGICISFGICNIGVAI 87
 QY 106 VDMVDSNFTLEDNRTSKACPEHSAPIKVHHNQTGRKKYQDAETQGMILGSFFGYIITQI 165
 Db 88 VSMVNNSTT-----HRGCHVYVQAKQ---FNMDPEVGLIHGSFEGVITVQI 132
 QY 166 PGGYVASKIGKMLGFGILGTAULTFTPIADLVGPLYVIRALEGIEGVTPPAMA 225
 Db 133 PGGFICQKTPAANRVFGFAIVATSTLNMILPSAARVHYGCVIFRILQGLVEGYTPACHG 192
 QY 226 MMSWAPLERKSLISTYAGQDGTIVSLPSIGIICYYMNMTYVFPFGTIGIFMFLM 285
 Db 193 IWSKNAPLERSLATTATPCGSYAGAVVAMPPLAGVLYVQSGHSVYVYVGSFGITFLRW 252
 QY 286 IWLVSPTPOKHKRISHYEKEYLISLRNQLSSQSY---PWVPIKSLPLMAIVVAH 340
 Db 253 LTVSYESPALHPISSEERKYIEDAIGESAKLNPVTKFNTPRRPFSTMPYAITVIANF 312
 QY 341 SYMTFTYLLTLPTYMKELIRPNVOENGLSLPYLGSMLCMLISGQADNIRAKMNS 400
 Db 313 CRSMFTYLLTLISQPAVFEVFEIGVLSKGLVSLPHLVMITIVIPIGQIADFLSRHMS 372
 QY 401 TILCVRIEFLIGMIGPAVFLVAAAGTIGDYSLAVAFITSTTIGCGSSGFSINHLDIAP 460
 Db 373 TTNVRKLMNCGRGMAITLLVYGY-SHKGVAISFLVLAVGSGAIGSFNNHLDIAP 431
 QY 461 SYAGILGINTFATIPGAVPIAKSLTPDNTVGEMQVIFYIAAINVEGAFITFLPAK 520
 Db 432 RYASILMISNGVGLSGMVCPIIVGAMTKRKREMOVYFLASLVHGVIFVCPAS 491
 QY 521 GEVQVMA 527
 Db 492 GEKQPMVA 498

RESULT 4
 A56410
 sodium/phosphate transport protein, renal - rabbit
 N:Alternate names: sodium/phosphate cotransporter
 C:Species: *Oryctolagus cuniculus* (domestic rabbit)
 C:Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 05-Nov-1999
 C:Accession: A56410; S27951
 R:Werner, A.; Moore, M.L.; Mantel, N.; Biber, J.; Semenza, G.; Murer, H.
 Proc. Natl. Acad. Sci. U.S.A. 88, 9608-9612, 1991
 A:Title: Cloning and expression of cDNA for a Na/P-1 cotransport system of kidney cor
 A:Reference number: A56410; MUID:92052140
 A:Accession: A56410
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-465 <WER>
 A:Cross-references: GB:W6466; NID:g165689; PIDN:AAA31461.1; PID:g165690
 C:Keywords: kidney; transmembrane protein

Query Match 29.0%; Score 822.5; DB 2; Length 465;
 Best Local Similarity 37.4%; Pred. No. 9,7e-58;
 Matches 172; Conservative 83; Mismatches 196; Indels 9; Gaps 4;

| | | | |
|----|-----|---|--|
| Qy | 74 | PVCSAAYNLTALIALEPFFIVYALRVNLSEVALUDWDNSTLTEDNRTSKACENHAPKY | 133 |
| | | 11 | PCFSEFRVYALNHEHFCNIVYIAQRKCLSTYVAMV-NNTNLHG5-----PNTSAERKL |
| Db | 11 | | 63 |
| Qy | 134 | HHNQTGRKYOMDAETOGMILGSEFFUYITITOLRGVASKIGKMLGGLGTAVLTF | 193 |
| Db | 64 | -DNKNRVYMMSPVQOITFSSIFEGAFILQIPVGYISISYIKKLIGALEFLSSVIF | 122 |
| Qy | 194 | TPRIADIGVGBLYLRALBESLGSEVTPRANHAMSSWAPDLERSKILSTYAGAOGLTVI | 253 |
| Db | 123 | IPQAAAGETWIIYICRVVQOITGTVTQAHETWYMAPRLRGRLTSMLSGFLGPFI | 182 |
| Qy | 254 | SLPISIIICYYMMNTYVYEFYEGFTIGTFWELTWLWISDPRKHRLSHYEREYILISRN | 313 |
| Db | 183 | VLYLTGIIICESLGNPWFYIFGACGCAVCLMFLVLYDDPKHPCVSLHEKEYITISLIQ | 242 |
| Qy | 314 | QLSS-QKSVFWPLKLSLPMAITVAHFSYNNTFYTLTLTPTYKELLRFVNOENGFLS | 372 |
| Db | 243 | QGSSTROSLPTKAMIKRILPMALISFCCFAYLMTYSRLIYPTFLINSMHVDIRENGLLS | 302 |
| Qy | 373 | SLPISGWLCLMILSGAADMLRAKNNFSLVCRIRFISLGLMIGRAVFLVYAGFICDLSL | 432 |
| Db | 303 | SLPFLPAMICGVANGTADFLMSRNNLSLTAIRKLTALGALLPFLVFSKCLLYLESGEYS | 362 |
| Qy | 433 | AVAFILISTLTGFCSSGEFINHLDAIPSYAGILGLITFTFATIGGOMPVYIAKSLTPDN | 492 |
| Db | 363 | TIPTLLIANNSSSPCLGGALINMLDLAPRYVFIKGVTLTLMIGTGMSSVYAGFLSDQ | 422 |
| Qy | 493 | TVGEMOTVFYIAAALNVFCAIFFTLPAKEGVOMALNDH | 532 |
| Db | 423 | PESSMEFIFILMSTINIVSYIFLIPAKKEIDMAEKOH | 462 |

```

RESULT      5
S69915
sodium-phosphate transport system 1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999
C:Accession: S69915
R:Chong, S.S.; Korak, C.A.; Liu, L.; Kristjansson, K.; Dunn, S.T.; Bourdeau, J.E.; Hughes,
Am. J. Physiol. 268, 1038-1045, 1995
A:Title: Cloning, genetic mapping, and expression analysis of a mouse renal sodium-depen
A:Reference number: S69915
A:Accession: S69915
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-465 <CHO>
A:Cross-references: EMBL:X77241; NID:G887522; PIDN:CAA54459.1; PID:G887523

```

[illegible][illegible]

```

RESULT      6
T43650
probable sodium-dependent inorganic phosphate cotransporter - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 09-Jun-2000 #sequence: revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: T43650
R:Lee, R.Y.N.; Sawin, E.R.; Chalfie, M.; Horvitz, H.R.; Avery, L.
submitted to the EMBL Data Library, September 1998
A:Description: Ear-4, a homolog of a mammalian sodium-dependent inorganic phosphate c
A:Reference number: Z22599
A:Accession: T43650
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-563 <LE>
A:Cross-references: EMBL:AF095787; PIDN:AAC64972.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: ear-4
A:Map position: 3
C:Function:
A:Description: is necessary for glutamatergic neurotransmission

```

```

Query Match          28.5%; Score 808; DB 2; Length 563;
Best local Similarity 36.4%; Fred. No. 1.7e-56;
Matches 166; Conservative 88; Mismatches 176; Indels 26; Gaps

OY 77 CSARNLALILFFGFEIYVALRVNLISVALVDVDSNTLEEDNRTSKACDEHSAPI-KVHN 135
   | : | | | | : : : | : : | : | | | | | | | | | | | | | | | | | | | | | |
Db 53 CRRMLLAILIAMGSMISGICRNGAAKTKHYKNYTD-----PRGKVMH 97
OY 136 NOTGKKYOMDATQGMILGSFFYGIITQIQEGGYASAKIGCKMLIGFGLTAUULTFPR 195
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 98 HE----FNNYTDILSVMESSYFGLYALQIPAGFLAAKPPKKLGFQFGVGAFINILRP 153
196 IAADGVGGLI-LRALDEIGEVYFRRHAMWSSMAPRLERSKILSISYAGAOIGTYS 254
154 YGFKXSDYLVAFIDITGLGVGVGYCPAMHGVRWRMAPRMESKILATTAFTGYSAGAVLG 213
OY 255 LPLSGIICYNNWTVYVFPEFTIGTFWRLMLWLVSDRPQKKRISHYEKET---LSSL 311
   | | | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 214 LPLSFLVSVYSWAPRFLYGVGYIMALMVCFFEEKRPAFRPTISOEKIPIEDAIGHV 273
312 RMLSSOKSVPPVPLIKSLPLMAIYVAHFSYMWTFYTLTLTPYMKELIRFVNOENGL 371
274 SNTHTTINSIPKALVTSKRPVAILVANSRBSWTFYLLQNDLTITMKELGKINDSGLL 333
OY 372 SSLPYLGSMLCILSGAANDNIRAKWNFTSLCVRIKFSLIGMIGPAVLVAAGTIGCDYS 431
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 334 AAIPLHVMGCVVLMGGOADYLRSKNKILSTTVVRKIFINFGGGGGAFAFLIAYATTS-D 392
432 LAVAFLLITSTLGGSCSSFINHDIAPSYAGILGITNTFATPIPGMWPIAKSLPPD 491
Db 393 TAIMLIAAVGSGRAISGFNNHDIAPRYAAILMGPSNGIGTLAGLCPVTEAFIAH 452
492 NTVGEMQVIFYIAAIIWFGAIEFTLLFAKGEVONNA 527
453 SKHG-WTSVFLLASLIHFTGVYFVAVVYSGEIOEWA 487

```

```

RESULT
7
H88548
protein ZK512.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: H88548
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: H88548
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1576 <SMO>
A:Cross-references: GB:chr_III; PIDD:CAA60150.1; PID:g338I690; GSPDB:GN00021; CESP:ZK512.6
C:Genetics:
A:Gene: ZK512.6
::Map position: 3

```

[illegible]

```

RESULT      8
A48916
sodium phosphate transport protein 1, renal - human
C:Species: Homo sapiens (man)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Nov-1999
C:Accession: A48916
R:Chong, S.S.; Kristjansson, K.; Zoghbi, H.Y.; Hughes, M.R.
Genomics 18, 355-359, 1993
A:Title: Molecular cloning of the cDNA encoding a human renal sodium phosphate transporter
A:Reference number: A48916; MUID:94117004
A:Accession: A48916
:Status: preliminary

```

A:Molecule type: mRNA
A:Residues: 1-467 <CHO>
A:Cross-references: GB:W1355; NID:g450531; PIDN:CA50490.1; PID:g450532
C:Gene(s):
A:Gene: GDB:SLC17A1, NP11
A:Cross-references: GDB:I41899, OMIM:182308
A:Map position: 6p23-6p21.3

| Query Match | Similarity | 28.3% | Score | 401.5 | DB | 2 | Length | 467 |
|-------------|------------|---|-------|-------------|-----|--------|--------|------|
| Best Local | Similarity | 38.2% | Pred | No. 4.6e-56 | | | | |
| Matches | 179 | Conservative | 78 | Mismatches | 185 | Indels | 27 | Gaps |
| QY | 74 | PVCSGARYNLALIAEFFEYIVYALRYNLVSVALVDVDSNTT--LEDNRTSKACPEHSADI | 131 | | | | | |
| DB | 13 | PGCFSEFRYSLEFLVACCNVITFAORACCLTLVWVNNNSDTPGLPTSTKCLYLDINIKNM | 72 | | | | | |
| QY | 132 | KVHNHQTGKKYOWDADEQSMIIIGSEFFGYIITQIPGGYVASKIGGKMIIIGFGLGTAVLT | 191 | | | | | |
| DB | 73 | -----YMSPDIGQITLISSTSYGIIITQVPYGISYTRKKMIGFALCLSSVUS | 122 | | | | | |
| QY | 192 | LFTPIAADLGVLIVLRALLEGIGVTFPRAMHAMWSNAPRLERSKLLISISYAGAQLT | 251 | | | | | |
| DB | 123 | LTIIPRAGIGVAMVYVCRAVQCAAGIYVTAQFEIYVKAAPLERGLRSLTSMSTSGFLCP | 182 | | | | | |
| QY | 252 | VTSPLSGIICCYMMWYTYFFFGYTGIIIGFWLLMTLMSVSDTPQKKRISHYKEYTLSSL | 311 | | | | | |
| DB | 183 | FTVLVLTGTIGCESLGNPMVFYIFGACGCAVCLLMFVLEFYDDKDRHPCISISKEYITSSL | 242 | | | | | |
| QY | 312 | RNQL-SQSKSVPMVWILKSLPLMAIVVAHFYSNMTFYTLTLTPYMKELIFENVOENF | 370 | | | | | |
| DB | 243 | VQGVSSRSQSLPIKALIKSLPYMALISIGSEFFFWNSHNMTEVTPEFINSMLHVNIKENGF | 302 | | | | | |
| QY | 371 | LSSLPLGSMCLMISGQADNLRKAKMNFSTLCVRRISFLISMIGPAPFLVAAGFTGCDY | 430 | | | | | |
| DB | 303 | LSSLPLPAMICGNLAGQISDFELTRNIIISVAVKRLFLPAAGFLLPAIFGVCLPALSSTF | 362 | | | | | |
| QY | 431 | SLAVAFLTSTTLGGFCSSGFSINHLDIAPSYAGILGTLTTFATIPGAVGVIAKSLT- | 489 | | | | | |
| DB | 363 | YSIVIEFLIAGTSGFCIGGVTFINGDLIAPRFEGFI---KACSLTGMIGGLINSTLTG | 418 | | | | | |
| QY | 490 | -----PDNTVGEMQTVFIYAAAINVFCALIFTLTAKGCVQWMLNDHH | 532 | | | | | |
| DB | 419 | LTIKODPESA---WEKTFILMAAINVTGTFELFIVATAIADIOAMAEKQH | 464 | | | | | |

RESULT 9
T23589
hypothetical protein K10G9.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T23589; T24636
R:MoreInfo, B.
submitted to the EMBL Data Library, August 1994
A:Reference number: Z19765
A:Accession: T23589
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-573 <W11>
A:Cross-references: EMBL:Z235282; PIDN:CAA85289.1; GSPDB:GN00021; CESP:K10G9.1
A:Experimental source: clone K10G9
R:Buick, D.
submitted to the EMBL Data Library, February 1995
A:Reference number: Z19915
A:Accession: T24636
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-573 <W12>
A:Cross-references: EMBL:Z48055; PIDN:CAA8135.1; GSPDB:GN00021; CESP:K10G9.1
A:Experimental source: clone T07A5
C:Genetics:
A:Gene: CESP:K10G9.1
A:Map position: 3

DB 121 LLIPAAIGVAMVVCRAVQGAQGI VATAQFEIYKMAPPLERGLTISMSTSGFLLGP 180

Db 380 FVRDPVIAVTCLVIACTGSGSVLSGFNVNHFEDIAPRYAPILMGIANGLGAVAG-VGGMVT 438

438

QY 486 KSLFEDNTVGEWOTVFYIAAIVNGAIFFTLPAKGEYONMA 527
 Db 439 NTVTYQNPBG-WKWPFLMAAIDIFGVIFFLIFAKGVLPMWA 479

RESULT 12

H84698
 hypothetical protein At2g29650 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: H84698
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shee, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 Euse, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402:761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: H84698
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-512 <STO>
 A:Cross-references: GB:AE002093; NID:93582333; PIDN:AAC3230.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g29650
 A:Map position: 2

Query Match 26.3%; Score 747; DB 2; Length 512;
 Best Local Similarity 35.1%; Pred. No. 1,1e-51;
 Matches 157; Conservative 80; Mismatches 166; Indels 44; Gaps 7;

QY 80 RYNLAIAFFGFYIVYALRVNLVALVDMVDSNTTLEDNRTSKACPEHSAPIKVHHNQTG 139
 Db 102 RNVIVLFCFSFLCNMDRVMSIALIPM-----S 131
 QY 140 KRYQMDAETQGMILGSEFYGIITQIPGCVASKIGKMLGFIGLTAVALFTLPPIAAD 199
 Db 132 AEYGMNPATVGLIQSFEWGYLLQIAGIWDVFGKRVLFGVIMWSIATILTPIYAK 191
 QY 200 LGVGPVLYRLBELGSEVTPPRAHAMSSAPPLERSKLLISYAGAQLTGVSPLSG 259
 Db 192 LGLPYLLVRAFKGVGEVAVAPRANMNLISKVWPQVRSRSLALVSGMVTGSVGLAFSP 251
 QY 260 IICYMMNTYVFYFGTIGIFWFLMLVLDTPQKHRIHYKEYLLSSLRNLSQK 319
 Db 252 FLIHQFGPSTVFYFGSGITWMLTLKASPLDEPTLPERKLLADMCASK-EPVK 310
 QY 320 SVPVPIILSLPLMAIVVAHFSTWTEYTLTLPTWKETLRFNVQENGFLSLPYLGS 379
 Db 311 SIPRLLISKPPVALLSCHCHNWGTFILLTWPPTYVHQVLFKFMESGLSVFP--- 366
 QY 380 WLCMILSGQA---ADNLRKWNSTLCVRIFSLIGIGPAPVLAAGFICGQYSLAVA 435
 Db 367 WMTMAISANAGMTADTLVSR-GFSVTVNRKIMOTIGLGPAPFLQLKHDSPTMAVL 424
 QY 436 FLITSTLGGCCSGGFSINHDIAPSYAGILGINTFATIPGAVPIAKSLTPDNTVG 495
 Db 425 CMACSGQGDATASQSGIVSNHODIAPRYSGLVGLSNTAGVLAGVLTATGHIIOH---G 481
 QY 496 EMQTVFYIAAIVNGAIFFTLPAKGE 522
 Db 482 SMDDVFITSVGLVGVITVNLVSTGE 508

RESULT 13
 B89135
 protein F25G6.7 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
 C:Accession: B89135
 R:anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
 A:Reference number: A75000; MUID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
 A:Accession: B89135
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-501 <STO>
 A:Cross-references: GB:chr_V; PIDN:AMC25800.1; PID:g2384837; GSPDB:GN00023; CESP:F25G
 C:Genetics:
 A:Gene: F25G6.7
 A:Map position: 5

Query Match 22.2%; Score 629.5; DB 2; Length 501;
 Best Local Similarity 31.6%; Pred. No. 2.5e-42;
 Matches 153; Conservative 101; Mismatches 191; Indels 39; Gaps 14;

QY 62 PLTPGAPRAEAPVCC---SARYNLIAFFGFYIVYALRVNLVALVDMVDSNTTLED 117
 Db 6 PISDGP-SQORPLTYRFPSPWRMTSIMLFCGCVHLMNSNMGMIVCMVNSSATY-D 63
 QY 118 NRTSKACPEHSAPIKVHHNQTGKRYQMDAETQGMILGSEFYGIITQIPGCVASKIGK 177
 Db 64 NET---YPENAAPL-----LDWSSDQGYIFSAFAGLLVMLETGG-MADKRNK 109
 QY 178 MLTGFGILGTAVLFL--FT-PIADLGVPLIVRALEGEGVTPPRAHAMSSAPPL 234
 Db 110 YMT---LVSVAIASLANFTLPRMAPISYVAIFSRFLVGRADLLOPRAMSLITRWFPTS 166
 QY 235 ERSKLLISYAGAQLTGVSPLSGIIC---YYMNTYVFYFGTIGIFWFLMLVLS 290
 Db 167 ERSYALGLANGGRQIGTLIIIPAGALCSOTEIFGWPISFYLSGFGVLFISYIFLGA 226
 QY 291 DTPQKHRIHYKEYLLSSLRNQ---LSQKSVPVPIILSLPLMAIVVAHFSTWTF 346
 Db 227 DKPSQKSCISNELKFTTISNOSQEDVGKKRERKVPKNKILSGAIVASVLSVCHPEPL 286
 QY 347 YTLTTLPTWKETLRFNVQENGFLSLPYLGSWLCMILSGQADNLR-AKNMSTLCV 404
 Db 287 MTLNMFPLSYLHDVHNHSTENGILSLPYSLMFAIGSSYLNTWLOKMTWKKDITC- 345
 QY 405 RRIIFSLGIGPAPVLAAGFICGQYS-LAVAFITSTLGGCCSGGFSINHDIAPSYA 463
 Db 346 KVLNLSIGSIGLIGFLAAATFLDNEHMAVLEFLCLSMASAGLTPCCQALAVSVAPAYS 404
 QY 464 GILGINTFATIPGMGPVIAKSLFPDNTVGEWOTVFYIAAIVNGAIFFTLPAKGEV 523
 Db 405 GAVTGFTEFFVAVSGIIPITTKIVDKRTAAEMNLVFIYSTVIAIPPIIIFNVWGSTEY 464
 QY 524 QMWA 527
 Db 465 QMWA 468

RESULT 14
 T01534
 hypothetical protein A.IG005110.nm - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 17-Mar-2000
 C:Accession: T01534
 R:Andrews, S.
 submitted to the EMBL Data Library, July 1997
 A:Description: The sequence of A. thaliana IG005110.
 A:Reference number: Z14347
 A:Accession: T01534
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-413 <AND>
 A:Cross-references: GB:AF013293; NID:g2252823; PID:g2252847
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 4

A: Introns: 79/3; 115/3; 128/3; 200/3; 288/3; 331/3; 353/3; 382/1
A: Note: A_IG005110.nm
C: Superfamily: hexuronate transporter

| Query Match | Score | DB | Length |
|-------------|-------|----|--------|
| 19.88 | 562 | 2 | 413 |

[illegible]

RESULT 15
 T45634
 hypothetical protein FL112.30 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
 C:Accession: T45634
 R:Choline, N.; Robert, C.; Brotlier, P.; Wincker, P.; Catolico, L.; Artiguenave, F.; S
 submitted to the Protein Sequence Database, November 1999
 A:Reference number: Z23010
 A:Accession: T45634
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-537 <CHO>
 A:Cross-references: EMBL:AL133292
 A:Experimental source: cultivar Columbia; BAC clone FL112
 C:Genetics:
 A:Map position: 3
 A:Introns: 35/2; 153/3; 229/2; 350/3; 399/3; 445/3; 457/1
 A:Note: FL112.30

| | | | | |
|-----------------------|--------------|--------------------|-----------------|-------------|
| Query Match | 18.7%; | Score 531; | DB 2; | Length 537; |
| Best Local Similarity | 27.6%; | Pred. NO. 1.9e-34; | | |
| Matches 151; | Conservative | 86; | Mismatches 195; | Indels 116; |
| | | | | Gaps 15; |

```

QY      33 ASTAHVGVMS-----PVDDLARN-----DGESTORTPLLEAPR 69
      20 SSSSLGCVIYSSFDWRIPELLPERDRRKLVLCTGRVYNSLKTGNTSVD---LCGIPR 75
QY      70 AEAAPVCCSARYN-----LAIIAFGFFIYALRVNLSVA 104
      11

```

| | | | |
|----|-----|---|-----|
| Db | 76 | HLRLVSCDARRPEETAELTAQPNFSEFITSERVAVVAMLA-LALALCNADRVMSVA | 134 |
| QY | 105 | LYDVAWDSTLTLEDNRTSKACEHSAPLKVHNNOTGKRYQMDAETQGNILGSFFGIYITQ | 164 |
| Db | 135 | IV-----PLLSRG-----WSKSFSGIVQSSFFLWGYLISP | 164 |
| QY | 165 | IPGGVYASKIGKKMLGHLGTAVLTLETPFIADLVGPFLIVRALEGEGEVTPEAMH | 224 |
| Db | 165 | IAGTGLVDRYGKRYVMAGVALMSLAFITLPWADSSLMLLARAAVGAEGVALPCMN | 224 |
| QY | 225 | AMSSMAPIERSKLLSISVAGQGLTVISLPLSGIICYYNNMVTYVFFPGTIGIFEWLL | 284 |
| Db | 225 | NMVARNPEPTERRSAVGIAMAGFOLGNVVGIMLSPILMSOGGIGPVIIGLSGFTMLLV | 284 |
| QY | 285 | WMLVSDPKOKHRISHXEKEYI-----LSLRNQLSSQSKV--PWVPIILKSLPLMAIVVA | 338 |
| Db | 285 | WLSTASSAPDRHPDITISELEYIQOKQOISTEMENKRIISTGCIPEGRILSKMPTAVYA | 344 |
| QY | 339 | HFSTYNNFTYLLLTLLPTVMEKILFENVOENGFLSSLPYLGSLMCLMISGOAA--DNLRA | 395 |
| Db | 345 | NSMSHWYHV-----NLKQAAMFSAVP-----WSMAATPFGIAGWSPDLI | 385 |
| QY | 396 | KWNSTLCVRRIEFLSLIGICPAVLVAAGFIGOYSL-AVAFLIISTLIGPQSSGSPIN | 454 |
| Db | 386 | RKRGSTILTRKIMOSIGFIGPGLI--GLTTAKQPLVASNMISLAAGLSFSHGLGIN | 443 |
| QY | 455 | HLDIAPSYAGILLIGITNTEAFIPGMVPIYAKSLTPDNTVAGEMQVTEFYIAAIVNFGAIF | 514 |
| Db | 444 | LOETIAPEYSGLVHMCMTAGTLAIVGTGAGFEV--ELLGSFGFTLLTAIYLLLSALF | 501 |
| QY | 515 | FTLEAKGE 522 | |
| Db | 502 | YNTATGE 509 | |

Search completed: July 15, 2002, 08:22:58
Job time: 43 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 08:22:15 : Search time 17.92 seconds
(without alignments)
1158.129 Million cell updates/sec

Title: US-09-776-865-2

Perfect score: 2836
Sequence: 1 MAAGAMTPRPVQPARPGF.....LPAKGEVQWALNDHGHHRH 536

Scoring table: BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | ID | Description |
|------------|-------|--------------------|------------|--------------------|
| 1 | 987 | 34.8 | YLD2_CAEEL | Q03567 caenorhabdi |
| 2 | 822.5 | 29.0 | NPT1_RABIT | Q28722 o renal sod |
| 3 | 809.5 | 28.5 | NPT1_MOUSE | Q61983 mus musculi |
| 4 | 808 | 28.5 | Y006_CAEEL | P36444 caenorhabdi |
| 5 | 806.5 | 28.4 | NPT1_RAT | Q62795 rattus norv |
| 6 | 801.5 | 28.3 | NPT1_HUMAN | Q14916 homo sapien |
| 7 | 767.5 | 27.1 | YRT3_CAEEL | Q10046 caenorhabdi |
| 8 | 686 | 24.2 | NPT3_HUMAN | Q00624 homo sapien |
| 9 | 615.5 | 21.7 | NPT4_HUMAN | Q00476 homo sapien |
| 10 | 486.5 | 17.2 | YKH4_CAEEL | P34272 caenorhabdi |
| 11 | 360 | 12.7 | GUDP_BACSU | P42237 bacillus su |
| 12 | 349 | 12.3 | GUDP_ECOLI | Q46916 escherichia |
| 13 | 339.5 | 12.0 | DGOT_ECOLI | P31457 escherichia |
| 14 | 326 | 11.5 | YRBO_BACSU | P37489 bacillus su |
| 15 | 318.5 | 11.2 | EXUT_BACSU | Q34456 bacillus su |
| 16 | 310 | 10.9 | GARP_ECOLI | P42613 escherichia |
| 17 | 270 | 9.5 | EXUT_ECOLI | P74470 escherichia |
| 18 | 267 | 9.4 | YFAV_ECOLI | P74470 escherichia |
| 19 | 264 | 9.3 | YJUL_ECOLI | P38398 escherichia |
| 20 | 258.5 | 9.1 | GUDP_ESEPU | P42205 pseudomonas |
| 21 | 242 | 8.5 | TUB3_AGRVI | P70786 agrobacteri |
| 22 | 236 | 8.3 | TUB4_AGRVI | Q44470 agrobacteri |
| 23 | 216 | 7.6 | YB04_HAEIN | P71369 haemophilus |
| 24 | 198 | 7.0 | PHT1_PSEPU | Q05181 pseudomonas |
| 25 | 194.5 | 6.9 | GLPT_BACSU | P37948 bacillus su |
| 26 | 190 | 6.7 | GLPT_HUMAN | P57057 homo sapien |
| 27 | 189.5 | 6.7 | GLPT_ECOLI | P08194 escherichia |
| 28 | 172.5 | 6.1 | UHPT_CHLMU | P08194 chlamydia m |
| 29 | 169.5 | 6.0 | YWFA_BACSU | P33637 bacillus su |
| 30 | 168.5 | 5.9 | YFET_ECOLI | P77228 escherichia |
| 31 | 168 | 5.9 | YJHB_ECOLI | P33352 escherichia |
| 32 | 168 | 5.9 | MUCK_ACICA | P94131 actinobact |
| 33 | 167.5 | 5.9 | YAOI_SCHPO | Q10097 schizosacch |

| | | | | | | |
|----|-------|-----|-----|---|------------|--------------------|
| 34 | 166 | 5.9 | 457 | 1 | PCAK_ACICA | Q43975 actinobact |
| 35 | 166 | 5.9 | 480 | 1 | GLPT_HAEIN | P96335 haemophilus |
| 36 | 164.5 | 5.8 | 455 | 1 | UHPT_CHLPN | Q94799 chlamydia p |
| 37 | 163 | 5.7 | 486 | 1 | CMCT_NOCIA | Q04723 nocardia la |
| 38 | 161 | 5.7 | 496 | 1 | NANT_ECOLI | P41036 escherichia |
| 39 | 160.5 | 5.7 | 403 | 1 | MHPT_ECOLI | P77589 escherichia |
| 40 | 160.5 | 5.7 | 515 | 1 | YMT2_RAT | Q01827 rattus norv |
| 41 | 160 | 5.6 | 451 | 1 | YFAJ_BACSU | P37514 bacillus su |
| 42 | 160 | 5.6 | 543 | 1 | SGEL_YEAST | P33335 saccharomyc |
| 43 | 156.5 | 5.5 | 464 | 1 | GALP_ECOLI | P37021 escherichia |
| 44 | 154.5 | 5.4 | 388 | 1 | NORA_STAM | P21191 staphylococ |
| 45 | 154.5 | 5.4 | 396 | 1 | BCK_ECOLI | P28246 escherichia |

ALIGNMENTS

| RESULT | ID | YLD2_CAEEL | STANDARD: | PRT: | 493 AA. |
|--------|--|------------|-----------|------|------------|
| AC | 003567: | | | | |
| DT | 01-FEB-1994 (Rel. 28, Created) | | | | |
| DT | 01-NOV-1997 (Rel. 35, Last sequence update) | | | | |
| DT | 01-MAR-2002 (Rel. 41, Last annotation update) | | | | |
| DE | Hypothetical 52.7 kDa protein C38C10.2 in chromosome III. | | | | |
| GN | C38C10.2. | | | | |
| OS | Caenorhabditis elegans. | | | | |
| OC | Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; | | | | |
| OC | Rhabditidae; Pelodierinae; Caenorhabditis. | | | | |
| OX | NCBI_TaxID=6239; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=BRISTOL N2; | | | | |
| RX | MEDLINE=94150718; PubMed=7906398; | | | | |
| RA | Wilson R., Almscough R., Anderson K., Baynes C., Berks M., Coulson A., | | | | |
| RA | Bonfield J., Burton J., Connell M., Copey T., Cooper J., Fraser A., | | | | |
| RA | Croxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A., | | | | |
| RA | Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., | | | | |
| RA | Johnston L., Jones M., Kershaw J., Kirsten J., Laister N., | | | | |
| RA | Lattelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M., | | | | |
| RA | Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R., | | | | |
| RA | Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden K., | | | | |
| RA | Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., | | | | |
| RA | Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J., | | | | |
| RA | Wohlman P.; | | | | |
| RT | *2.2 Mb of contiguous nucleotide sequence from chromosome III of C. | | | | |
| RT | elegans.* | | | | |
| RL | Nature 368:32-38(1994). | | | | |
| RN | [2] | | | | |
| RP | REVISIONS. | | | | |
| RA | Jones S.J.M.; | | | | |
| RA | Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases. | | | | |
| CC | -1- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1. | | | | |
| CC | -1- This SWISS-PROT entry is copyright. It is produced through a collaboration | | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | | |
| CC | or send an email to license@sib-sib.ch). | | | | |
| CC | ----- | | | | |
| DR | EMBL: 219153; CAA79549.1; - | | | | |
| DR | PIR: S28286; S28286. | | | | |
| DR | WormPep: C38C10.2; CE08647. | | | | |
| DR | InterPro: IPR003662; sub-transporter. | | | | |
| DR | Pfam: PF00083; sugar tr. 1. | | | | |
| KW | Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport; | | | | |
| KW | Sodium transport. | | | | |
| FT | TRANSMEM 10 30 | | | | POTENTIAL. |
| FT | TRANSMEM 85 105 | | | | POTENTIAL. |
| FT | TRANSMEM 112 132 | | | | POTENTIAL. |


```

RESULT 3
ID NP1L_MOUSE STANDARD: PRT: 465 AA.
AC 061983:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Renal sodium-dependent phosphate transporter protein 1 (Sodium/phosphate
DE cotransporter 1) (Na(+)/Pi cotransporter 1) (renal sodium-phosphate
DE transport protein 1) (Renal Na(+)-dependent phosphate cotransporter
DE 1).
GN SLC12A1 OR NP1L.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=9535846; PubMed=7611445;
RA Chong S.S., Kozak G.A., Liu L., Kristjansson K., Dunn S.T.,
RA Bourdeau J.E., Hughes M.R.; and expression analysis of a mouse renal
RA "Cloning, genetic mapping, and expression analysis of a mouse renal
RA sodium-dependent phosphate cotransporter."
RL Am. J. Physiol. 268:F1038-F1045(1995).
CC -1- FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY.
CC MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA
CC Na+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: KIDNEY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X77241; CA54459.1; -.
DR MCD: MGI:103209; Slc12a1.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr: 1.
KW Transport; Transmembrane; Glycoprotein; Symport; Sodium transport.
FT TRANSSEM 79 99 POTENTIAL.
FT TRANSSEM 117 137 POTENTIAL.
FT TRANSSEM 176 196 POTENTIAL.
FT TRANSSEM 199 219 POTENTIAL.
FT TRANSSEM 260 280 POTENTIAL.
FT TRANSSEM 304 324 POTENTIAL.
FT TRANSSEM 337 356 POTENTIAL.
FT TRANSSEM 363 383 POTENTIAL.
FT TRANSSEM 399 419 POTENTIAL.
FT TRANSSEM 429 449 POTENTIAL.
FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 465 AA; 51589 MW; C67EE25A2C291EEF CRC64;

```

```

Query Match 28.5%; Score 809.5; DB 1; Length 465;
Best Local Similarity 36.7%; Pred. No. 8.4e-52;
Matches 167; Conservative 90; Mismatches 189; Indels 9; Gaps 2;

```

```

OY 74 PYCCSARNALAILAFECFFIYALRVNLVALVDMVDSNTLTEDNRFSKACPEHSAPIKV 133
DB 11 PGFCFGRGLAILLHFCINAIAMQRCVCLNTLMVAV-----NNTGSPHLSESSEYVM 62
OY 134 HHNQTGKYQDAETQGTILSPFYGIITQIPGYSVASKTGKMLGFGILGTAVALTLF 193
DB 63 LDNKNPNYSPDIQGLILSSVFEGMAYVAPVGYLSGITPMKRITGSSLFLSLMSLL 122

```

```

OY 194 TPRIADLCVGLIVLRALEGVTEPPAMHAMSSNAPPLERSKLISVYAGLGTVI 253
DB 123 IPPAAOVGVALVIVCRVLOGIAQGVSTGGEIWMKAPPLERGLTSMTLISGFVGPFI 182
OY 254 SLPLSLGICCYMNTYVFEFGTIGIEFLLMVLVSTPQKHRIKSHYEYILSSRN 313
DB 183 VLVSGFCIDLLGPMPVEYIFGIVGCVLISWFLFEDDPKDHDPWMSSEKDYIISLPMO 242
OY 314 QLS-OKSPVWPLIKSLPLMAIVAHVSVMYTYTLTLTPTWKLRLRNVQENPLS 372
DB 243 QASSGRSLPIKAMKSLPLMAITLINSFATWSNLSLVTYPTISTVLYHVNENGLS 302
OY 373 SLPLYSWLCMLILSGOADNLRANKNESTLCVRRIFSLIGMIGPAVPLVAGFGCDVSL 432
DB 303 SLPLILAVICILLAGQMSDFLTRKISIVYVRKLTLLSGFCVPIFMCLLYSYNYS 362
OY 433 AVAPLTISTYLGRCSSGFSFNHLDIAPSVAGILGINTFATIPGMVPIYAKSLPDPN 492
DB 363 TVIFELTANSTLFSYSCQLINALDIAPRYGCFKAVATLIGMFGILSTLGLILMOD 422
OY 493 TVGEGQTFEYIAALINVGALFETLFAKEGYONNA 527
DB 423 PEYAMHKISFLMAGINTCLVFLEFLFAKGLQDMA 457

```

```

RESULT 4
ID Y006_CAEEL STANDARD: PRT: 576 AA.
AC P34644;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical 52.2 kDa protein ZK512.6 in chromosome III.
GN ZK512.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favallo A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkes T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z22177; CA80150.1; -.
DR PIR: S40767;
DR Wormpep: ZK512.6; CE01109.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr: 1.
KW Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport;

```



```

RESULT 6
NPT1_HUMAN
ID NPT1_HUMAN STANDARD: PRT; 465 AA.
AC 014916; 013783;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Renal sodium-dependent phosphate transporter protein 1 (Sodium/phosphate
DE cotransporter 1) (Na(+)/PI cotransporter 1) (Renal sodium-phosphate
DE transport protein 1) (Renal Na(+)-dependent phosphate cotransporter 1)
DE (NA/PI-4).
GN SLC17A1 OR NPT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94117004; PubMed=8288239;
RA Chong S.S., Kristjansson K., Zoghbi H.Y., Hughes M.R.;
RT "Molecular cloning of the cDNA encoding a human renal sodium
RT phosphate transporter protein and its assignment to chromosome
RT 6p21.3-p23."
RL Genomics 18:355-359(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RX MEDLINE=95126933; PubMed=7826357;
RA Miyamoto K.-I., Tatsumi S., Sonoda T., Yamamoto H., Minami H.,
RA Taketani Y., Takeda E.;
RT "Cloning and functional expression of a Na(+)-dependent phosphate co-
RT transporter from human kidney: cDNA cloning and functional
RT expression."
RL Biochem. J. 305:81-85(1995).
CC -!- FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY.
CC MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA
CC NA+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN KIDNEY CORTEX, LIVER AND BRAIN
CC BUT NOT IN OTHER TISSUES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X71355; CAA50490.1; ALT_INIT.
DR EMBL: D28532; BAA05888.1; -.
DR MIM: 182308;
DR InterPro: IPR003662; sub-transporter.
DR Pfam: PF00083; sugar_tr. 1.
KW Transport; Transmembrane; Glycoprotein; Symport; Sodium transport.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 176 196 POTENTIAL.
FT TRANSMEM 198 218 POTENTIAL.
FT TRANSMEM 255 275 POTENTIAL.
FT TRANSMEM 299 319 POTENTIAL.
FT TRANSMEM 337 357 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 431 451 POTENTIAL.
FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 35 36 RA -> LM (IN REF. 2).
FT CONFLICT 77 77 I -> V (IN REF. 2).
FT CONFLICT 207 207 G -> C (IN REF. 2).

```

```

FT CONFLICT 231 231 S -> G (IN REF. 2).
SQ SEQUENCE 465 AA; 50884 MW; 687A19F245C4ED16 CRC64;
Query Match 28.3%; Score 801.5; DB 1; Length 465;
Best Local Similarity 38.2%; Pred. No. 3.2e-51;
Matches 179; Conservative 78; Mismatches 185; Indels 27; Gaps 6;
OY 74 PVOCSARYKLATIAFGFFIYVALRVNLSVALVDWDSMT--LEENRFSKACPEHSAPI 131
DB 11 PGFCSPRGLSLVHCNNVITRQACLNLTVMVWNSDHPGLPTSTKKLIDNIKNP 70
OY 132 KVHNQTKKRYOMDAETQGMILSGSFYGIITQIPGQYASKIGKMLGFGIITAVLT 191
DB 71 -----YKMSPIQGIILSTSYGIIITQVPYSGIYSTKMKMGFALCLSSVLS 120
OY 192 LFTPIADVGEPILYRALBEGVGTPPAHAMSSNAPPLERSKLLSISYAGQLOT 251
DB 121 LLIPTAGIGVAMVVVCRAVQAGIVATAQFEIYVKMAPLERGLRSLMSYSGFLGP 180
OY 252 VISLPLSGIICYYMNTVYFEPFGTIGIFMPLIMLVSDPQKHRIHYEKYITLSL 311
DB 181 FVLVLTGVICSLGMPWYIFGACGACVCLIMFVLYDDPKDPCISISEKEYITSSL 240
OY 312 RNDL-SSOKSVPVPLIKSLPLMAIVAAHSYWNFTYLTLLPTMYKEILRFNVQENF 370
DB 241 VQGVSSRSQSLPIKALIKLPLWALISGSPFPWSNINITYTPMINSMLHNINENF 300
OY 371 LSLPLYSMLCMLISGQADNLRAKWNFSYLCVRRIISLIGNIGDAVLVAAGTICDY 430
DB 301 LSLPLPLFAMICGNLAGQLSDFELTRNIIISVIAVRKLFPAAGFLPAIGVCILPYLSSTF 360
OY 431 SLAVAFILITSLTGFCSSGFSTNHDIAPSYAGILGTINFPATIRGWPVPIANSLT 489
DB 361 YSTVIFLLIAGATGSCFLGCVFLNGIDIAPIRIFGLT-----KACSTLIGIGLITSLTG 416
OY 490 -----PDNTVGEMQVTFYIAAINVFGAIFFLFAKGHVQWMLNDH 532
DB 417 LLIKQPEEA---WFKTFILMAINVTGLIFYLIVATAIQDAKRNKH 462
RESULT 7
YRT3_CAEBL
ID YRT3_CAEBL STANDARD: PRT; 544 AA.
AC 010046;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Hypothetical 60.5 kDa protein T07A5.3 in chromosome III.
DE T07A5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitoidea;
OC Rhabdilitidae; Felodertinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Buck D.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: 248055; CAA88134.1; -.
DR WormPeP: T07A5.3; CE01648.
KW Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport;

```



```

OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RL "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP POSSIBLE FUNCTION.
RX MEDLINE=98447507; PubMed=97722162;
RA Hubbard B.K., Koch M., Palmer D.R., Babbitt P.C., Gerlt J.A.;
RT "Evolution of enzymatic activities in the enolase superfamily:
RT characterization of the (D)-glucarate/galactarate catabolic pathway
RT in Escherichia coli.";
RL Biochemistry 37:14369-14375(1998).
RN [3]
RP GENE NAME.
RX MEDLINE=20225875; PubMed=10762278;
RA Monterrubio R., Baldoma L., Obradors N., Aguilar J., Badia J.;
RT "A common regulator for the operons encoding the enzymes involved in
RT D-galactarate, D-glucarate, and D-glycerate utilization in
RT Escherichia coli.";
RL J. Bacteriol. 182:2672-2674(2000).
CC -1- FUNCTION: UPTAKE OF D-GLUCARATE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE PHTHALATE PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U29581; AAA0439.1; -
DR EMBL: AE00362; AAC75831.1; -.
DR Ecocore: EG13169; gndp.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
KW Transmembrane; Transport; Complete proteome.
FT TRANSMEM 21 41
FT TRANSMEM 59 79
FT TRANSMEM 98 118
FT TRANSMEM 154 174
FT TRANSMEM 177 197
FT TRANSMEM 255 275
FT TRANSMEM 291 311
FT TRANSMEM 330 350
FT TRANSMEM 352 372
FT TRANSMEM 388 408
FT TRANSMEM 417 437
FT TRANSMEM 450 491
SO SEQUENCE 450 AA; 49142 MW; E0599FACB1BB488 CRC64;

Query Match 12.3%; Score 349; DB 1; Length 450;
Best Local Similarity 24.5%; Pred. No. 2.7e-18;
Matches 115; Conservative 90; Mismatches 197; Indels 68; Gaps 16;

QY 78 SARYNALIAAFGEFFIVLALRYNLVALYMDVSNLTLDNRKSKCPHSAPIKYNHNO 137
DB 16 NARYWLVVWLFYTSFNVDRAATLSAGSEM----- 46
QY 138 TGRKYQMDAETQWILGSEFFYGIITQIPGVYASKIGKMLGILGTAVLT--FT 194
DB 47 -ANDIGLDVGMKYVSASMAVIGQIRGWLIDRFSGKRYVFWSEIFLWMTLLQGV 105
QY 195 PIAADIG-VGRLIVRALBGLGCVTPPMNHAMWSSWADPLERSKLLISYACAGOLGTVI 253

```

```

DB 106 DIFSGFGIIVALFTLRFVLGLAEDSPFGNSRIVAAMPDAQERGTAVSIFNSAQIFATVI 165
QY 254 SPLSGIICYYNMWTVYEFYFGTIGFELTWIMLVSDTPQKHKRISHREKEYI-----L 308
DB 166 FAPIQWMLTHEGWSHVFPEFGKGLIVISFIMLKATHE-PNQHPCVKNKELEIYAAGNL 224
QY 309 SGLRNQLSSQK--SVWPVI--LKSLEPLMAIVAHFSYNTFYTLTLLEPTYKEILR 362
DB 225 IMMDQNTKVKVPFSVKMQGQIKQLLGRSMIGVYIGQYICINLTFEFTIWPFPVYLQANG 284
QY 363 FVQENGFLSLPYLGSWICMLLSQADNLAKNWFSLCYRRIFSL-IGMIGPAV--F 419
DB 285 MSILKAGFVASYPAVCGFVGVLGGIISD-----W-----LMRRRTGSLNIAKRTPIVMGM 334
QY 420 IYAAGRIGCDYS---LAVAFILISTTGGFCSSGSINHLIAP-SYAGIILGINTPFA 474
DB 335 LLSMVAWPCNYVNVEMWIIIFMALAFEGGIGALGMAY-MADTAPREISGLSGGLNMF 393
QY 475 TIPGMVPIYAKSLTPDNTVGEWQVIFYIAAINVFGALFTLFAKGEVO 524
DB 394 NSGIYTPAIAGIYV--GTTGSFNGAL-IYGVHIALIVLSYLVLDGIR 440

RESULT 13
ID DQOT_ECOLI STANDARD; PRT; 430 AA.
AC P31457;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE D-galactonate transporter.
GN DQOT OR B3691.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=93315143; PubMed=7686882;
RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication.";
RN [2]
RP SIMILARITY TO DAL5 FAMILY.
RA Koonin E.V.;
RL Unpublished observations (OCT-1993).
CC -1- FUNCTION: INTAKE OF GALACTONATE INTO THE CELL.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
CC -----
CC -1- SIMILARITY: BELONGS TO THE PHTHALATE PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L10328; AAA62043.1; ALT_INIT.
DR EMBL: AE000446; AAC76714.1; ALT_INIT.
DR Ecocore: EG11715; dgot.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
KW Transmembrane; Inner membrane; Transport; Complete proteome.
FT TRANSMEM 18 38
FT TRANSMEM 55 75
FT TRANSMEM 85 105
FT TRANSMEM 143 163
FT TRANSMEM 168 188
FT TRANSMEM 243 263

```

FT TRANSMEM 284 304 POTENTIAL.
 FT TRANSMEM 310 330 POTENTIAL.
 FT TRANSMEM 338 358 POTENTIAL.
 FT TRANSMEM 376 396 POTENTIAL.
 FT TRANSMEM 404 424 POTENTIAL.
 SQ SEQUENCE 430 AA; 47076 MW; 86B1902777F2E461 CRC64;

Query Match 12.0%; Score 339.5; DB 1; Length 430;
 Best Local Similarity 26.3%; Pred. No. 1,3e-17;
 Matches 124; Conservative 79; Mismatches 185; Indels 83; Gaps 18;

QY 80 RYNDLALAFGEFTVYALRVNLSVALVDVDSNTLEDNRTSKACPEHSAPIKVHNQGT 139
 DB 14 RYLLVLMIFIVVICYDRAANLAVA-----SAHIOEFGIT- 49
 QY 140 KKYQMDAETOGMILGSEFFGYIIIPGGYVASKIGCKMLLFGILGTAVLTLPITAD 199
 DB 50 -----KAE-MGYFSARAWLYTLCOIPGWFLDKRVGSVYVFIAIFGMSVATLFOGFRAT- 102
 QY 200 LGVGPLIVLRALGELGEGVTFPRAMHMSWAPPLERSKLTLSYAGAOLGTVISLPLSG 259
 DB 103 -GLMSLIGLRITGIFEARAPFTNNRVTSWFERERASANGFTSGOFVLAFLTPLL 161
 QY 260 IICYMMNTVYFFFGTIGIFWLLIMVSDPQKHKRISHYKEYI----- 307
 DB 162 WQEMLSMHWFIVTGGIITMSLIM-FKYQRPRLTKGISKALDLYIRDGGLVDGDPAR 220
 QY 308 -LSSLRNQDSQKSVPPVPIKSLPRMAIYVAHFSYNNFTLTLLTPYKELLRNVQ 366
 DB 221 VKKRAEPLTAK--DW-KLVEHRRKLGVYIGQFAVASTLMEFLTPENNYLTOKGITAL 276
 QY 367 ENGELSLPYGSLWMLSCGADNLRANKWFSTLCVRRIFSL-IGMIGPAV--FLVAA 423
 DB 277 KAGMTVTPFLAAVGVLSGVAD-----LVYKRGFSIGPARKPIITGCLLIST 326
 QY 424 GFIGCDYS-----LAVAFLITSTLGFCSSGFSINHLDIAP-SYAGILGITTTF 473
 DB 327 CIMGANTYNDPMIMICMLALAFEG-----NGFASITWSLVS-SLAPARKLIGTGVNEFA 380
 QY 474 ATIGMGVPYAKSLTPDNTVGEWQVFTAAAINVFGALFTFLPAKGEVQ 524
 DB 381 GGLGIGTIVPLVGYLA--QGEGFAPALVYI-SAVALLGALSTILLY-GDVK 427

RESULT 14
 YRBO_BACSU
 ID YRBO_BACSU STANDARD; PRT; 435 AA.
 AC P37489;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical transport protein yrbO.
 GN YRBO.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96051385; PubMed=7584024;
 RA Ogasawara N., Nakai S., Yoshikawa H.;
 RT "Systematic sequencing of the 180 kilobase region of the Bacillus
 subtilis chromosome containing the replication origin.";
 RL DNA Res. 1:1-14(1994).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE PHALATE PERMEASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; D26185; BAA05188.1; -;
 DR EMBL; Z99124; CAB16094.1; -;
 DR Subtilist; BG10016; yrbO.
 DR Interpro; IPR003662; sub_transporter.
 DR Pfam; PF00083; sugar_tr; 1.
 KW Hypothetical protein; Transmembrane; Transport; Complete proteome.

FT TRANSMEM 26 46 POTENTIAL.
 FT TRANSMEM 61 81 POTENTIAL.
 FT TRANSMEM 96 116 POTENTIAL.
 FT TRANSMEM 119 139 POTENTIAL.
 FT TRANSMEM 150 170 POTENTIAL.
 FT TRANSMEM 177 197 POTENTIAL.
 FT TRANSMEM 242 262 POTENTIAL.
 FT TRANSMEM 281 301 POTENTIAL.
 FT TRANSMEM 325 345 POTENTIAL.
 FT TRANSMEM 347 367 POTENTIAL.
 FT TRANSMEM 385 405 POTENTIAL.
 FT TRANSMEM 407 427 POTENTIAL.
 SQ SEQUENCE 435 AA; 48248 MW; 9630914D9B606208 CRC64;

Query Match 11.5%; Score 326; DB 1; Length 435;
 Best Local Similarity 25.7%; Pred. No. 1.2e-16;
 Matches 116; Conservative 78; Mismatches 193; Indels 64; Gaps 16;

QY 90 GFPIV-YALRVNLSVALVDVDSNTLEDNRTSKACPEHSAPIKVHNQGTQYQMDAET 148
 DB 29 GIITLNFDRVAISVAAPAIQDS-----FHLTATE 58
 QY 149 QGWLISSEFFGYIITQIPGGYVASKIGCKMLLFGILGTAVLTLPITADLGVPLIV 208
 DB 59 LGIVFSITVYTYLTMQPVGSLDLDRGVAVWTRVGMTWSFLLTL-LAPLOGKLLLYLF 116
 QY 209 RALEGLGEGVTFPRAMHMSWAPPLERSKLTLSYAGAOLGTVISLPLSGIICYNNWT 268
 DB 117 RFLIGLTSASAFPAASKATATLMPPSPRGLANSIFDSAAFSNVIGAPLVAFLVTFDWK 176
 QY 269 YVEFFFGTIGIFWLLIMVSDPQKHKRISHYKEYILSLRNQDSQKSVPPV--PT 326
 DB 177 VAFETIGICINVL-FTIFPMQYEQPERHKKRISSELYI--QKHNATITBQ-IPYKGPL 232
 QY 327 LKSL-----PLMAIYVAHFSYNNFTLTLLTPYKELLRNVQENGFSLSLPYGSLWC 382
 DB 233 LKSLFTNRKRWGLMIGFTGYGFNLLTWLPFFKHTYGMDLMSGLFAVFWLLISTIS 292
 QY 383 MILSGQADNLRANKWFSTLCVRRIFSLIGMIGPRAVFLVAGFTGCDYSIAVAFITISTY 442
 DB 293 GIANGGLVQVFIKKGYPNKKVRTVIVGMSGFFL---GSI-LTNNTVAMICISIG 348
 QY 443 LGGFCSS---GFSINHLDIAP---SYAGILGITTTFATIPMGVGPYAKSLTP--DNT 493
 DB 349 LAGISATAPVGSWS-AELPAGISVSMLSMVNLAN-----NLFGIIAASLTGLYLFV 401
 QY 494 VGEWQVFTAAAINVFGALFTFLPAKGEVQ 524
 DB 402 TGSFTLSFLVAGFVLLGLVGY-VFVLGDVYK 431

RESULT 15
 EXUT_BACSU
 ID EXUT_BACSU STANDARD; PRT; 422 AA.
 AC O34456;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hexurionate transporter.
 GN EXUT.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;

alignment_block:
us-09-776-865-2 x AA250879

Align seg 1/1 to: AA250879 from: 1 to: 2930

```

1  MetaLaalAGlyAlaMetThrProProArgProValGlnProAlaArgPr 17
   |||||
263 ATGGCGGGGGGGGATGACACCGCCCGCGGTCAGCAGCTCGGCC 312
   |||||
17  oGlyGlyPheGlyLeuSerGlyArgArgSerLeuLeuGlyValAlaAs 34
   |||||
313 CGGGGGCTTCGGGCTGTCGGGCGGCTCCCTCTCTGCGCAGGtGGCGA 362
   |||||
34  eRThrProAlaHisValGlyValMetArgSerProValArgAspLeuAla 50
   |||||
363 GTACACCTGCTCAGCTAGCGCTCATGAGTCTCCGGTTCGAGACCTGGCC 412
   |||||
51  ArgAsnAspGlyGluGluSerThrAspArgThrProLeuLeuProGlyAl 67
   |||||
413 CGGAACGATGGCGAGAGACGACGACCGCACGCTCTTCTACCGGGCGC 462
   |||||
67  aProArgIaGluAlaAlaProValCysCysSerAlaArgTyrAsnLeuA 84
   |||||
463 CCCAGGGGCGGAGCGCTCCAGTGTGCTGCTGCTGTACAACTTAG 512
   |||||
84  IaIleLeuAlaPhePheGlyPhePheIleValTyrAlaLeuArgValAsn 100
   |||||
513 CAATTTGGCCTTTTGGTTGTTCTTCATGTGTATGCTACGCTGTAAT 562
   |||||
101 LeuSerValAlaLeuValAspMetValAspSerAsnThrThrLeuGluAs 117
   |||||
563 CTGAGTGTGGCTAGTGTAGTATGTAGATTCAAATACAACTTAGAAGA 612
   |||||
117  PaSaArgThrSerLysAlaCysProGluHisSerAlaProIleLysValH 134
   |||||
613 TAAATGAACCTTCCAAAGCGGTGCCAGACATCTGCTCCATAAAGATTC 662
   |||||
134  ISHisAsnGlnThrGlyLysLysTyrGlnTrpAspAlaGluThrGlnGly 150
   |||||
663 ATCATTAATCAAAAGCGGTAGAGAACATCGAATGGAGAAACCAAGGA 712
   |||||
151 TrpIleLeuGlySerPhePheTyrGlyTyrIleIleThrGlnIleProG 167
   |||||
713 TGGATTCCTCGGTTCCCTTTTATGCTACATCATCACAGATTTCCCGG 762
   |||||
167  YGlyTyrValAlaSerLysIleGlyLysMetLeuLeuGlyPheGlyI 184
   |||||
763 AGGATATGTGCCAGCAAAATAGGGGGGAAATGCTGCTAGGATTTGGGA 812
   |||||
184  IeLeuGlyThrAlaValLeuThrLeuPheThrProIleAlaAlaAspLeu 200
   |||||
813 TCCTTGGGACGCTGCTCCTCACCTGTTCACCTCCATTGCTGCAGATTTA 862
   |||||
201 GlyValGlyProLeuIleValLeuArgAlaLeuGluGlyLeuGly 217
   |||||
863 GGAGTTGGACCACTCATGTACTCAGAGCACTAGAGACTAGAGAGAGGG 912
   |||||
217  YValThrPheProAlaMetHisAlaMetTrpSerSerTrpAlaProPro 234
   |||||
913 TGTTCACATTTCCACGCATGATGATGATGTGTTCTTGGGCTCCCTC 962
   |||||
234  euGluArgSerLysLeuLeuSerLysSerTyrAlaGlyAlaGlnLeuGly 250
   |||||
963 TTGAAAGAGCAAACTTCTTAGCATTTTCGATTCGAGAGACACACTTGGG 1012
   |||||
251 ThrValIleSerLeuProLeuSerGlyIleIleCysTyrTyrMetAsnTr 267
   |||||
1013 ACAGTAATTTCTCTCTCTCTTCTGGAATATATTGCTACTATATGAATG 1062
   |||||
267  pThrTyrValPheTyrPhePheGlyThrIleGlyIlePheThrPheLeu 284
   |||||
1063 GACTTATGTCTTCTTACTTTTGTGTAATAATGGAATATTGTTGTTCTTT 1112

```

```

284  eutrIleTrpLeuValSerAspThrProGlnLysHisLysArgIleSer 300
   |||||
1113 TGTGATCTGTGTTAGTAGTACACACACCAAAACAAACAGAGATTTC 1162
   |||||
301 HisTyrGlnLysGlyTyrIleLeuSerSerLeuArgAsnGlnLeuSer 317
   |||||
1163 CATTAATGAAGAATATCATTTCTTTCATCATTAAGAATATCAGCTTCTTC 1212
   |||||
317  rGlnLysSerValProTrpValProIleLeuLysSerLeuProLeuTrp 334
   |||||
1213 ACAGAAGTCACTGCCGTGGTACCATTATTAATTCCTCGCACTTTGGG 1262
   |||||
334  IaIleValAlaIleHisPheSerTyrAsnTrpThrPheTyrThrLeuLeu 350
   |||||
1263 CTATCGTAGTTGCACACTTTTCTTACAACGTGACTTTTATACTTATATG 1312
   |||||
351 ThrLeuLeuProThrTyrMetLysGluIleLeuArgPheAsnValGlnG 367
   |||||
1313 ACATTATTTGCTCTACTTATATGAAGAGATCTTAAGCTTCAATGTTCAAGA 1362
   |||||
367  uAsnGlyPheLeuSerSerLeuProTyrLeuGlySerTrpLeuCysMet 384
   |||||
1363 GAATGGGTTTTATCTTCATTTGCTTATTTAGGCTTTGGTTATGATAGA 1412
   |||||
384  IeLeuSerGlyGlnAlaAlaAspAsnLeuArgAlaLysTrpAsnPheSer 400
   |||||
1413 TCGTGTCTGGTCAGAGCTGCTGACAAATTTAAGGCAGAAATGGAATTTTCA 1462
   |||||
401 ThrLeuCysValArgArgIlePheSerLeuIleGlyMetIleGlyProAl 417
   |||||
1463 ACTTTATGTGTCTCCAGAAATTTTAGCCTTATAGAAATGATTTGGACCTGC 1512
   |||||
417  aValPheLeuValAlaIaGlyPheIleGlyCysAspTyrSerLeuVal 434
   |||||
1513 AGTATTCCTGTAGCTGCTGCTGCTTATGCTGATTAATCTTTGGCCG 1562
   |||||
434  aAlaPheLeuThrIleSerThrThrLeuGlyGlyPheCysSerSerGly 450
   |||||
1563 TTGCTTTCCTACTATATNCAACAACTGGGAGGCTTTTCTGCTTCGGA 1612
   |||||
451 PheSerIleAsnHisLeuAspIleAlaProSerTyrAlaGlyIleLeu 467
   |||||
1613 TTTAGCATCAACCATCTGGAATATGCTCTTCGTATGCTGTATCCCTCT 1662
   |||||
467  uGlyIleThrAsnThrPheAlaThrIleProGlyMetValGlyProVal 484
   |||||
1663 GGGCATTCACAAATCATTTGGCACATATCCAGGAATGTTGGGCCCTCA 1712
   |||||
484  IeAlaLysSerLeuThrProAspAsnThrValGlyGluTrpGlnThrVal 500
   |||||
1713 TTGCTAAAGTCTGACCCCTGATTAACACTGTGGAGAAATGGCAACCGTG 1762
   |||||
501 PheTyrIleAlaIleAlaIleAsnValPheGlyAlaIlePhePheThrLe 517
   |||||
1763 TTTCTATATTTGCTGCTCTATTAATGTTTGTGTCCTATTTCTTTTACACT 1812
   |||||
517  upheAlaLysGlyGluValGlnAsnTrpAlaLeuAsnAspHisHisGlyH 534
   |||||
1813 ATTGCGCAAGGTGAAGTACAAACTGGGCTCTCAATGATCACCATGGAC 1862
   |||||
534  IsArgHis 536
   |||||
1863 ACAGACAC 1870

```

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAID10325

seq_documentation_block:

ID AAD10325 standard; DNA; 2930 BP.

XX

AC AAD10325;

XX

DT 16-OCT-2001 (first entry)

| | |
|---|--|
| XX | Human group B beta-haemolytic Streptococci toxin receptor (HP59) DNA. |
| XX | |
| KW | Human; group B beta-haemolytic Streptococci toxin receptor; GBS; HP59; |
| KM | cyclostatic; vulnereary; antiatherosclerotic; osteopathic; vasotrophic; |
| KV | wound prevention; attenuation; pathoangiogenic condition; cancer; scar; |
| KW | wound healing; gliosis; nerve injury; chronic wound; reperfusion injury; |
| KM | kelooid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis; |
| KM | vaccine; ds. |
| OS | Homo sapiens. |
| XX | |
| FH | Key |
| FT | Location/Qualifiers |
| CDS | 263..1873 |
| FT | /tag= a |
| FT | /product= "Human GBS toxin receptor protein (HP59)" |
| XX | WO200156598-A2. |
| PD | |
| XX | 09-AUG-2001. |
| PF | |
| XX | 02-FEB-2001; 2001WO-US03662. |
| XX | |
| PR | 02-FEB-2000; 2000US-0179870. |
| XX | |
| PA | (UYVA-) UNIV VANDERBILT. |
| XX | |
| PI | Hellerqvist CG; |
| DR | WPI: 2001-488844/53. |
| P-PSDB: | AAE06518. |
| XX | |
| PT | Preventing or attenuating pathoangiogenic conditions e.g. cancer, |
| PT | chronic wounds, osteoarthritis, kelooids and psoriasis in a mammal, by |
| PT | administering group B beta-haemolytic Streptococci toxin receptor or its |
| fragment - | |
| PS | Disclosure; Page 41-44; 52pp: English. |
| XX | |
| CC | The present sequence is a DNA encoding full length human group B beta- |
| CC | haemolytic Streptococci (GBS) toxin receptor protein, HP59. The present |
| CC | invention relates to a method for preventing or attenuating a patho- |
| CC | angiogenic condition in a mammal which comprises administering to the |
| CC | mammal one or more GBS toxin receptors or their immunogenic fragments |
| CC | to induce or maintain an immune response to one of GBS toxin receptors. |
| CC | The method is useful for preventing or ameliorating pathoangiogenic |
| CC | conditions such as cancer, scarring during wound healing, gliosis |
| CC | during repair of nerve injury, chronic wounds, kelooids, reperfusion |
| CC | injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and |
| CC | psoriasis in mammals. The proteins of the invention are also used |
| CC | as vaccines. |
| XX | |
| SO | Sequence 2930 BP; 729 A; 698 C; 681 G; 822 T; 0 other; |
| XX | |
| alignment_scores: | |
| Quality: 2836.00 | Length: 536 |
| Ratio: 5.291 | Gaps: 0 |
| Percent Similarity: 100.000 | Percent Identity: 100.000 |
| alignment_block: | |
| US-09-776-865-2 x AAD10325 .. | |
| Align seg 1/1 to: AAD10325 from: 1 to: 2930 | |
| 1 Metalaalaglyalametrrhprioproargprovalglnprolaarrpr 17 | |
| | |
| 263 ATGGCGGCGGGGCGATGACGCCGCCCGGTCAGCAGCTCGGCC 312 | |
| | |
| 17 ogllyllyheglyleuserglyaragarserleuleucysglnvalalals 34 | |
| | |
| 313 GGGGGGCTTGGGGCGGTGGGGCGGCGCTCTCTTGCCAGAGTGGCA 362 | |

| | | |
|------|--|------|
| 34 | erThProAlaHisValGlyValMetArgSerProValArgAspLeuAla | 50 |
| 363 | GTACACCTGCTCACCTAGCGCTGCATGAGAGTCTCCGGTTCGAGACCTGGCC | 412 |
| 51 | ATGAsnAspGlyGlnGlnSerThrAspArgPheProLeuLeuProGlyVal | 67 |
| 413 | CGGACGATGGGAGAGAGCGACGAGACCGCACGCGCTCTTCTACCGGGGCG | 462 |
| 67 | APrOArGAlaGlnAlaAlaProValCysCysSerLaaArgTyrAsnLeuA | 84 |
| 463 | CCGAGGGCGGAAGCGCGCTCCAGTGTGCTGCTGCTGTTACAACTTAG | 512 |
| 84 | IsrLeuAlaPhePheGlyPhePheIleValTyrAlaLeuArgValAsn | 100 |
| 513 | CAATTTTGGCGCTTTTGGTTCTTTCATTTGTGTATGACTTACGTGTGAAT | 562 |
| 101 | LeuSerValAlaLeuValAspMetValAspSerAsnTrpPheLeuGlnAs | 117 |
| 563 | CGAGCGTGGCTGTAGTGATAGTGAATGTAATTCAAATTAACCTTATGAAGA | 612 |
| 117 | PaSnArGTrhSerLysAlaCysProGlnHisSerLapProLeuLysValH | 134 |
| 613 | TATATGAACTTCCAGAGCGGTGTCCGAGACATTTCTGCTCCATAAAGTTC | 662 |
| 134 | IsHIsAsnGlnThrGlyLysLysTyrGlnTrpAspAlaGluThrGlnGly | 150 |
| 663 | ANCATATATCAACGGGTAAAGATGCACATGGAGATGCAGAAATCTCAAGA | 712 |
| 151 | TrpIleLeuGlySerPhePheTyrGlyTrpIleIleThrGlnIleProG | 167 |
| 713 | TGGATTTCTCGGTCTCTTTTATGTGCATACATCAACACAGATTCCTGG | 762 |
| 167 | NgLyIyrAlaAlaSerLysIleGlyGlyMetLeuLeuGlyPheGlyI | 184 |
| 763 | AGGATATGTGTGCACACAAAATAGGGGGGAAAAATGCTGTAGGATTTGGGA | 812 |
| 184 | IeLeuGlyThrAlaValIeuthrLeuPheTrpProIleAlaAlaAspLeu | 200 |
| 813 | TGCTTGGGACGCTGCTGCTCCACACCGTTCACCTCCATTTGGTGCAGATTTA | 862 |
| 201 | GlyValGlyProLeuIleValLeuArgAlaLeuGlnGlyLeuGlyGluI | 217 |
| 863 | GGAGTTGGACCACTATGTACTGTACAGACACTAGAAAGACATGAGAGAGGG | 912 |
| 217 | YValThrPheProAlaMetHisAlaMetTrpSerSerTrpAlaProProL | 234 |
| 913 | TGTTCATTTCCAGCGCATGATGCATGTGGTCTTCTTGGGCTCCCGCTC | 962 |
| 234 | eugLuarSerLysLeuLeuSerIleSerTyrAlaGlyAlaGlnLeuGly | 250 |
| 963 | TTGAAAGAAGCAAACTGTTAGCATTTGGTATGGAGAGGACAGACCTTGGG | 1012 |
| 251 | ThrValIleSerLeuProLeuSerGlyIleIleCysTyrTyrMetAsnTr | 267 |
| 1013 | ACAGTAATTTCTCTCCCTTCCTTCGAGATTAATTTGCTATATAGCAATTG | 1062 |
| 267 | PheTrpYrAlaPheTyrPhePheGlyThrIleGlyIlePheTrpPheLeu | 284 |
| 1063 | GACTTATGCTTCTTCACTTTTGGTACTATTGAGATTAATTTGGTTCTTT | 1112 |
| 284 | eutTrpIleTrpLeuValSerAspTrhProGlnLysHisLysArgIleSer | 300 |
| 1113 | TGTGATGTCGGTTAGTTAGTGCACACCAACCAAAAAACAAAGAAATTCC | 1162 |
| 301 | HisTyrGlnLysGluTyrTlleLeuSerSerLeuArgAsnGlnLeuSerSe | 317 |
| 1163 | CATTTATGAAAAGAAATACATTTCTTTCATCATTAAGAAATCAGCTTTCTTC | 1212 |
| 317 | rgLInLysSerValProTrpValProIleLeuLysSerLeuProLeuTrpA | 334 |
| 1213 | ACAGAAAGTCAAGCGCGTGGTGAACCATTTTAAATTCCTCGACATTTGGG | 1262 |
| 334 | IaIleValValAlaHisPheSerTyrAsnTrpPheTyrThrLeuLeu | 350 |

```
|||||
1263 CTATGGATGGACACCTTTCTTACAACTGACCTTTTAATCTTATG 1312
351 ThrLeuLeuProThrTyrMetLysGluLeuArgPheAsnValGlnG 367
|||||
1313 ACATTATGCGCTTATATATGAGAGATCCCAAGGTTCAATGTTCAAGA 1362
367 uAsnGlyPheLeuSerSerLeuProTyrLeuGlySerTrpLeuCysMet 384
|||||
1363 GAATGGGTTTTATCTCATTCCTTATTTAGGCTCTGGTATATGTATAGA 1412
384 LeuLeuSerGlyGlnAlaAlaAspAsnLeuArgAlaLysTrpAsnPheSer 400
|||||
1413 TCCTGTCTGGTCAAGCTGCTGACAAATTAAAGGCAAAATGAAATTTTCA 1462
401 ThrLeuCysValArgArgIlePheSerLeuIleGlyMetIleGlyProAl 417
|||||
1463 ACTTATATGTTGTCGAGAAATTTTAGCCTTATAGAAATGATGGACCTGC 1512
417 AlaIlePheLeuValAlaAlaGlyPheIleGlyCysAspTyrSerLeuAlaVal 434
1513 AGTATTCCTGGTAGCTGCTGGCTTCATTTGCTGATTTATCTTTGGCCG 1562
434 AlaIlePheLeuThrIleSerThrThrLeuGlyGlyPheCysSerSerGly 450
1563 TTGCTTCTCTACATATATCAACACACACGAGGCTTTGCTCTCTGCA 1612
451 PheSerIleAsnHisLeuAspIleAlaIleProSerTyrAlaGlyIleLeuLe 467
|||||
1613 TTTACATCAACCAATCGATATATTCCTCTCGATGCTGATCTCTCTCT 1662
467 uGlyIleThrAsnThrPheAlaThrIleProGlyMetValGlyProValI 484
1663 GGGCATCAACAATACATTTGCCACATTTCCAGAAATGGTGGCCCGCA 1712
484 IleAlaLysSerLeuThrProAspAsnThrValGlyIleTrpGlnThrVal 500
1713 TTGCTAAAGTCTGACCCCTGATTAACCTGTTGGAGAAATGGCAACCGTGG 1762
501 PheTyrIleAlaAlaAlaIleAsnValPheGlyAlaIlePhePheThrIle 517
1763 TTCATATATGCTGCTGCTATTAATGTTTGGTGCATTTCTTACTACT 1812
517 uPheAlaLysGlyGluValGlnAsnTrpAlaLeuAsnAspHisGlyH 534
1813 ATTCCCAAGGTAAGTACAAAACTGGGCTCTCAATGATCACCAATGAC 1862
534 IsArgHis 536
|||||
1863 ACAGACAC 1870
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAF55900
seq_documentation_block:
ID   AAF55900 standard; DNA; 2512 BP.
XX   AAF55900;
XX   AAF55900;
XX   18-APR-2001 (first entry)
XX   Human AST coding sequence.
XX   Human; AST; noctropic; immunotrophic; gene therapy; Salla disease;
XX   anion and sugar transporter; anion-cation symporter;
XX   sialic acid transporter; ss.
OS   Homo sapiens.
XX   EP1069184-A1.
XX   17-JAN-2001.
XX   16-JUL-1999; 99EP-0202341.
```

```
XX   16-JUL-1999; 99EP-0202341.
PR   (ALKU ) AKZO NOBEL NV.
XX   WPI; 2001-193090/20.
XX   P-PSDB; AAB6967.
DR   New human transporter gene implicated in Salla disease and lysosomal
PT   sialic acid transport, useful in assays for identifying new drugs, or
PT   diagnosing sialic acid transport defects related to mutations in the
PT   transporter gene.
XX   Claim 2; Page 12-13; 20pp; English.
XX   The present sequence is the coding sequence for human Anion and Sugar
CC   Transporter (AST) protein. AST has significant homology with several
CC   members of the anion-cation symporter (ACS) family of transporters. AST
CC   is implicated in Salla disease, and is useful in screening assays for
CC   identifying new drugs. Compounds identified via AST screening is useful
CC   for preparing a pharmaceutical suitable as an activator or inhibitor of a
CC   sialic acid transporter protein. The pharmaceutical may be used in sialic
CC   acid associated diseases and CNS/immune related disorders.
XX   Sequence 2512 BP; 612 A; 585 C; 566 G; 749 T; 0 other;
```

```
alignment_scores:
Quality: 2804.00      Length: 535
Ratio: 5.261          Gaps: 1
Percent Similarity: 99.626      Percent Identity: 99.252
```

```
alignment_block:
US-09-776-865-2 x AAF55900
```

```
Align seg 1/1 to: AAF55900 from: 1 to: 2512
```

```
4 GlyAlaMetThr.....ProProArgProValGlnProAlaArgProG 18
|||||
151 GGGGGGGGTGTCATCGCCCCCGCCGCGGCCAGCCAGCTCGGCCGG 200
18 yGlyPheGlyLeuSerGlyArgArgSerLeuLeuCysGlnValAlaSer 35
201 GGGCTTCGGGCTGTCGGGCGCGCGCTCCCTCTCTGCGCAGGTGGGAGATA 250
35 hrProAlaHisValGlyValMetArgSerProValArgAspLeuAlaArg 51
251 CACCTGCTCACGTAGGCTCATGAGGTCCGGTTCCGAGACCTGGCCCGG 300
52 AsnAspGlyGluGluSerThrAspArgThrProLeuLeuProGlyAlaPr 68
301 AACGATGGCGAGGAGAGACGACGCCGACGCCCTCTTACAGGGGCCGCC 350
68 ArgAlaGluAlaAlaAlaProValCysCysSerAlaArgTyrAsnLeuAla 85
351 ACGGCGCGAAGCCGCTCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 400
85 IleLeuAlaPhePheGlyPhePheIleValTyrAlaLeuArgValAsnLeu 101
401 TTTTGGCCTTTTGGTTTCTTCATTTGCTATGCTATGCTATGCTATGCTATG 450
102 SerValAlaLeuValAlaPheValAlaPheSerAsnThrThrLeuGluAsp 118
451 AGTGTGGCTAGTAGATATGATGATCAATCAATCAATCAATCAATCAATCA 500
118 ArgThrSerLysAlaCysProGluHisSerAlaProIleLysValHis 135
501 TAGAACTTCCAAAGGCTGTCAGAGCATTTGCTCCATTAAGATTCATC 550
135 IsAsnGlnThrGlyLysLysTyrGlnTrpAspAlaGluThrGlnGlyTyr 151
551 ATAATCAACGGGTAGAGATGCAATGGGATGCAAGAACTCAAGAGATGG 600
```


XX
SQ Sequence 2712 BP; 677 A; 619 C; 619 G; 797 T; 0 other;

alignment_scores:

Quality: 2734.00 Length: 556
Ratio: 5.238 Gaps: 1
Percent similarity: 93.885 Percent identity: 93.885

alignment_block:
US-09-776-865-2 x AAK94876 ..

Align seg 1/1 to: AAK94876 from: 1 to: 2712

```
15 AlaArgProGlyGlyPheGlyLeuSerGlyArgSerLeuLeuGly 31
|||||
2 GCTCGGGCCGGGGCTTCGGGCTGTCGGCCGGCCCTCCCTCTCTGCCA 51
|||||
31 nValAlaSerThrProAlaHisValGlyValMetArgSerProValArg 48
|||||
52 GGTGGCGAGTACACCTGCTCAGCTAGGGGTATGAGGTCTCGGTTTCGAG 101
|||||
48 spleAlaArgAsnAspGlyGluSerThrAspArgThrProLeuLeu 64
|||||
102 ACCCTGGCCCGAAGCATGGCGAGAGACGACGCCGACGCCCTCTCTCA 151
|||||
65 ProGlyAlaProArgAlaGluAlaAlaProValGlyCysSerAlaArgTy 81
|||||
152 CCGGGCGCCCGCGGGCCGAGAACCGCTCAGTGTCTGCTGCTGCTTA 201
|||||
81 rAsnLeuAlaIleLeuAlaPhePheGlyPhePheIleValTyrAlaLeu 98
|||||
202 CAACTTAGCAATTTTGGCCTTTTGGTTTCTTCAATGTGATGATGATAC 251
|||||
98 rGlyAlaAsnLeuSerValAlaLeuValAspMetValAspSerAsnThrThr 114
|||||
252 GTGTAAATCTGAGTGTGCTGATGATGATGATGATGATGATGATGATGAT 301
|||||
115 LeuGluAspAsnArgThrSerTyrAlaCysProGluHisSerAlaProIle 131
|||||
302 TTAGAAAGATATAGAACTTCCAGGCGTGTCCAGAGCATTTCTGCTCCCAT 351
|||||
131 eLysValHisHisAsnGlnThrGlyLysTyrGlnThrAspAlaGluT 148
|||||
352 AAAAGTTCATCATATCAACGCGTAAAGAGATGCAATGGATGCAAGAA 401
|||||
148 hArgGlyTyrIleLeuGlySerPhePheTyrGlyTyrIleIleThrGln 164
|||||
402 CTCAGGAGATGATCTCGGTTCTCTTTTATGAGCTACATCATCACACAG 451
|||||
165 IleProGlyTyrValAlaSerSerIleGlyGlyLysMetLeuLeuG 181
|||||
452 ATTCCTGGAGATATGTGCGCAGCAAAATAGGGGGAAATGCTGCTAGG 501
|||||
181 yPheGlyIleLeuGlyThrAlaValLeuThrLeuPheThrProIleAla 198
|||||
502 ATTGGGATTCCTGGCAGCTGCTGCTCACCCTGTTCACTCCCATTTGCTG 551
|||||
198 IaAspLeuGlyValGlyProLeuIleValLeuArgAlaLeuGluLeu 214
|||||
552 CAGATTGTGGAGTTGGACCATCTGATCTAGAGCACTAGAAAGAGCTA 601
|||||
215 GlyGluGlyValIleThrPheProAlaMetHisAlaMetTrpSerSerTrpAl 231
|||||
602 GGAGAGGGGTTCATTTCCAGCCCATGCAATGATGATGATGATGATGATG 651
|||||
231 aProProLeuGluArgSerTyrLeuLeuSerIleSerTyrAlaGlyAlaG 248
|||||
652 TCCCTCTCTGAAAGCAAACTCTTACATTTCAATATGAGAGAGAC 701
|||||
248 ILeuGlyThrValIleSerLeuProLeuSerGlyIleIleCysTyrTyr 264
|||||
702 AGCTTGGAGAGATATTTCTCTCTCTGAAATATTTGCTACATAT 751
```

```
265 MetAsnTrpThrTyrValPheTyrPhePheGlyThrIleGlyIlePheTr 281
|||||
752 ATGAAATGAGCTTATGCTCTTCTACTTTTGGTACTATATGAAATTTTGG 801
|||||
281 pPheLeuLeuThrPileTrpLeuValSerAspThrProGlnLysHisLysA 298
|||||
802 GTTCTCTTTGGATCTGGTATGATGATGATGATGATGATGATGATGATG 851
|||||
298 rGlyIleSerHisTyrGluLysGluTyrIleLeuSerSerLeuAsnAsn... 313
|||||
852 GAATTTCCATTATGAAAGAAATACATCTTTCATCATTTAAGAAATACAG 901
|||||
313 ..... 313
902 GGTGAGTGCAGTACATGATCTTGGCTGCTACGACCTCCCTCTGCTC 951
|||||
314 .....G 314
952 CCGGGTCAAGCATCTCCACCTCAGCTACCAAGTAGTTGGAGACGAC 1001
|||||
314 ILeuSerSerGlnLysSerValProTrpValProIleLeuLysSerLeu 330
|||||
1002 AGCTTCTTCAAGAAATGACGTGCGGTGATACCATTTTAAATCCCTG 1051
|||||
331 ProLeuTrpAlaIleValAlaHisPheSerTyrAsnTrpThrPheTy 347
|||||
1052 CCACCTTGGGCTATGCTGATGCTGACACCTTTCTTCAACTGACATTTT 1101
|||||
347 rThrLeuLeuThrLeuLeuProThrTyrMetLysGluIleLeuArgPheA 364
|||||
1102 TACTTATTTGACATATATGCTACTTATATGAAAGATTCCTAAGGTCA 1151
|||||
364 snValGlnGluAsnGlyPheLeuSerSerLeuProTyrLeuGlySerTrp 380
|||||
1152 ATGTTCAAGAGAAATGGTTTATCTTCATTTGATTTAGGCTCTTGG 1201
|||||
381 LeuCysMetIleLeuSerGlyGlnAlaAlaAspAsnLeuArgAlaLysTr 397
|||||
1202 TTATGTATGATCTCTGCTGCTCAGCTCTGACAAATTTAAAGGCAAAATG 1251
|||||
397 pAsnPheSerThrLeuCysValArgArgIlePheSerLeuIleGlyMetI 414
|||||
1252 GAATTTTCAACTTATGTGTCGCAAGAAATTTTAGCTTATAGGAATGA 1301
|||||
414 IleGlyProAlaValPheLeuValAlaAlaGlyPheIleGlyCysAspTyr 430
|||||
1302 TTGACCTGCAAGATATCTGTGAGCTGCTGCTTCAATTTGCTGATATAT 1351
|||||
431 SerLeuAlaValAlaPheLeuThrIleSerThrThrLeuGlyGlyPheCy 447
|||||
1352 TCTTTGGCCGCTTCTTCTTAACATATATCAACACACACGAGGCTTTTG 1401
|||||
447 sSerSerGlyPheSerIleAsnHisLeuAspIleAlaProSerTyrAlaG 464
|||||
1402 CTCCTTGTGATTTAGCATCAACCATCTGATATGCTCTCTGATGCTG 1451
|||||
464 IyIleLeuLeuGlyIleThrAsnThrPheAlaThrIleProGlyMetVal 480
|||||
1452 GTATCTCTCTGGGATCAAAATACATTTGCACTATTTCCAGGAATGTT 1501
|||||
481 GlyProValIleAlaLysSerLeuThrProAspAsnThrValGlyGluTr 497
|||||
1502 GGGCCCGTCATTTGCTAAAGTCTGACCCCTATTAACACTGTGGAGAAATG 1551
|||||
497 pGlnThrValPheTyrIleAlaAlaAlaIleAsnValPheGlyAlaIleP 514
|||||
1552 GCAAAACGCTGTATATATGCTGCTGATTAATGTTTGTGTCATTT 1601
|||||
514 hePheThrLeuPheAlaLysGlyValGlnAsnTrpAlaLeuAsnAsp 530
|||||
1602 TCTTTACATATATGCGCAAAAGTGAAGTAAACATGCGGCTCTCAATGAT 1651
```

```

531 HISHISGLYHISARGHIS 536
      |||||
1652 CACCATGGACACAGACAC 1669

```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AA250875

seq_documentation_block:
ID AAZ50875 standard; cDNA; 2602 BP.

AC AAZ50875;

DT 31-MAY-2000 (first entry)

DE Partial human GBS toxin receptor (HP55) cDNA.

KM Human GBS toxin receptor; group B beta-haemolytic streptococci; HP55.
KM pathological vascularisation; cancer metastases; angiogenesis;
KM neovascularisation; reperfusion injury; scarring; keloid;
KM chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KM endothelial cell proliferation; antibacterial; anticancer;
KM anti-angiogenic; anti-inflammatory; anti-arthritic; anti-psoriatic; ss

05 Homo sapiens.

| FH | Key | Location/Qualifiers |
|-----|-----|---------------------|
| 10 | 10 | 10 |
| 11 | 11 | 11 |
| 12 | 12 | 12 |
| 13 | 13 | 13 |
| 14 | 14 | 14 |
| 15 | 15 | 15 |
| 16 | 16 | 16 |
| 17 | 17 | 17 |
| 18 | 18 | 18 |
| 19 | 19 | 19 |
| 20 | 20 | 20 |
| 21 | 21 | 21 |
| 22 | 22 | 22 |
| 23 | 23 | 23 |
| 24 | 24 | 24 |
| 25 | 25 | 25 |
| 26 | 26 | 26 |
| 27 | 27 | 27 |
| 28 | 28 | 28 |
| 29 | 29 | 29 |
| 30 | 30 | 30 |
| 31 | 31 | 31 |
| 32 | 32 | 32 |
| 33 | 33 | 33 |
| 34 | 34 | 34 |
| 35 | 35 | 35 |
| 36 | 36 | 36 |
| 37 | 37 | 37 |
| 38 | 38 | 38 |
| 39 | 39 | 39 |
| 40 | 40 | 40 |
| 41 | 41 | 41 |
| 42 | 42 | 42 |
| 43 | 43 | 43 |
| 44 | 44 | 44 |
| 45 | 45 | 45 |
| 46 | 46 | 46 |
| 47 | 47 | 47 |
| 48 | 48 | 48 |
| 49 | 49 | 49 |
| 50 | 50 | 50 |
| 51 | 51 | 51 |
| 52 | 52 | 52 |
| 53 | 53 | 53 |
| 54 | 54 | 54 |
| 55 | 55 | 55 |
| 56 | 56 | 56 |
| 57 | 57 | 57 |
| 58 | 58 | 58 |
| 59 | 59 | 59 |
| 60 | 60 | 60 |
| 61 | 61 | 61 |
| 62 | 62 | 62 |
| 63 | 63 | 63 |
| 64 | 64 | 64 |
| 65 | 65 | 65 |
| 66 | 66 | 66 |
| 67 | 67 | 67 |
| 68 | 68 | 68 |
| 69 | 69 | 69 |
| 70 | 70 | 70 |
| 71 | 71 | 71 |
| 72 | 72 | 72 |
| 73 | 73 | 73 |
| 74 | 74 | 74 |
| 75 | 75 | 75 |
| 76 | 76 | 76 |
| 77 | 77 | 77 |
| 78 | 78 | 78 |
| 79 | 79 | 79 |
| 80 | 80 | 80 |
| 81 | 81 | 81 |
| 82 | 82 | 82 |
| 83 | 83 | 83 |
| 84 | 84 | 84 |
| 85 | 85 | 85 |
| 86 | 86 | 86 |
| 87 | 87 | 87 |
| 88 | 88 | 88 |
| 89 | 89 | 89 |
| 90 | 90 | 90 |
| 91 | 91 | 91 |
| 92 | 92 | 92 |
| 93 | 93 | 93 |
| 94 | 94 | 94 |
| 95 | 95 | 95 |
| 96 | 96 | 96 |
| 97 | 97 | 97 |
| 98 | 98 | 98 |
| 99 | 99 | 99 |
| 100 | 100 | 100 |

| | | |
|----|-----|----|
| FT | CDS | 58 |
|----|-----|----|

```
FT      /*tag= a
FT      /product= "partial human GBS toxin receptor"
```

PN W0200005375-A1.

PD 03-FEB-2000

PF 22-JUL-1999; 99WO-US16676.

PR 22-JUL-1998; 98US-0093843.

PA (UYVA-) UNIV VANDERBILT

PI HELLERQVIST CG, FU C;

DR WPI; 2000-205377/18.

XX

PT useful for diagnosis and treatment of, e.g. pneumonia in neonates -

PS Claim 3; Page 77-80; 109pp; English.

The present cDNA sequence encodes partial human GBS (group B beta-haemolytic streptococci) toxin receptor (HP25). This sequence was cloned by using human embryo lung cDNA library as template. Expression vectors comprising this cDNA can be transformed into host cells to express GBS toxin receptor and its fragments. Detecting the receptor in tissues is used to diagnose pathological vascularisation, e.g. for detecting cancer metastases. GBS toxin receptors are useful for treating conditions associated with pathological angiogenesis or neovascularisation (specifically cancer, reperfusion injury, scarring during wound healing, keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or neural injury), and to raise specific antibodies used for treating early onset disease. Inhibitors of this receptor are useful for treating pathological or hypoxia-induced endothelial cell proliferation and migration.

Sequence 2602 BP; 672 A; 576 C; 583 G; 771 T; 0 other;

alignment_scores:

Quality: 2717.00

Length: 514

Ratio: 5.286

Gaps:

similarity: 100.000

Percent Identity: 100.000

alignment_block:

US-09-776-865-2 x AAZ50875

Align seg 1/1 to: AAZ50875 from: 1 to: 2602

| | | |
|-----|--|-----|
| 23 | SerGlyArgArgSerLeuLeuGlyGlnValAlaSerThrProAlaHisVal | 39 |
| 1 | TCGGGGCGGGCGCCCTTCCTCCGACGGGTGGCGAGTACACCTGCTCAAGT | 50 |
| 39 | IGlyValMetArgSerProValArgAspLeuAlaArgsnAspIysGlnG | 56 |
| 51 | AGGCGTCATGAGGTCTCCGGTTCGAGACCTGGCCCGGACATGGCCAGG | 100 |
| 56 | IuSerThrAspArgThrProLeuLeuProGlyAlaProArgAlaGluAla | 72 |
| 101 | AGAGCAGCAGCAGCAGCAGCCTCTTTCACGGCGGCCCCACGGCCGAAGCC | 150 |
| 73 | AlaProValCysCysSerThrLArgTrpAsnLeuAlaIleLeuAlaPhePh | 89 |
| 151 | GGTCCAGGTGGCGGTCTGCTGCTGTACAACTTAGCAATTTTGGCTTTT | 200 |
| 89 | ecgIlePheAlaIleValArgAlaLeuArgValAsnLeuSerValAlaLeuV | 106 |
| 201 | TGGTTCTTCATGTGTGTATGCATTCAGTGTGATGACTGAGTGTCCGTAG | 250 |
| 106 | alAspMetValAspSerAsnThrThrLeuGlnAspAsnArgThrSerLys | 122 |
| 251 | TGGATTTGGTAGTTCAAATACACTTATAGAAGATTAATAGAACTTCCAG | 300 |
| 123 | AlCysProGlnHisSerLArgProIleLysValHisHisAsnGlnThrG | 139 |
| 301 | GCGGTGCCAGACGATCTTCCTCCATAAAAGTTCATCATTAATCAACGGG | 350 |
| 139 | yLysLysTrpGlnTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerP | 156 |
| 351 | TAAAGAAGTACCAATGGAGTGCAGAAACCAAGATGAGTCTCCGGTTCCT | 400 |
| 156 | hePheTrpGlyTrpIleIleThrGlnIleProGlyGlyTrpValAlaSer | 172 |
| 401 | TTTTTTATGGCTACATCAACACAGATTCCTGGAGGATTAATGTTGCCAGC | 450 |
| 173 | LysIleGlyGlyLysMetLeuLeuLysPheGlyIleLeuGlyThrAlaVal | 189 |
| 451 | AAAAATGGGGGGAATGCTGCTAGAGATTGGATCCCTGGCACTGCTGT | 500 |
| 189 | IleuThrIleuPheThrProIleAlaAlaAspLeuGlyAlaGlyProLeuI | 206 |
| 501 | CCCTACCCCTGTTCCACCCCATTTGCGAGATTATAGGAATTTGGACACTCA | 550 |
| 206 | IleValIleuArgAlaIleuGlnGlyLeuGlnGlyGlnGlyValThrPheProAla | 222 |
| 551 | TTGTACTCAGACGACTAGAGAGACTAGAGAGGGGTTCATCTTCCAGCC | 600 |
| 223 | MetHisAlaMetIrpSerSerThrPheProLeuGlnIuArgSerLysLe | 239 |
| 601 | ATGCATGSCCATGNGCTTCCTTGGGCTCCCTCTTGAAAGACCAAACT | 650 |
| 239 | IuSerIleIleSerTrpAlaGlyAlaGlnLeuGlyThrValIleSerIeuP | 256 |
| 651 | TCCTTACCATTTGCTATGCGAGGACACACTTGGGCACTAATTTTCTCTTC | 700 |
| 256 | roLeuSerGlyIleIleCysTrpTrpMetAsnTrpThrTrpValPheTrp | 272 |
| 701 | CTCTTTCGTGAATTAATTTCTCTATATGAAATGGACTTATGCTTCTTAC | 750 |
| 273 | PhePheGlyThrIleGlyIlePheThrPheLeuLeuThrIleTrpLeuVal | 289 |
| 289 | IleSerAspThrProGlnIleHisLysArgIleSerHisTrpGlnLysGlu | 306 |
| 751 | TTTTTTGGTACATATGGAATATATTGGTTCTTTTGGTGGATCGTGTAGT | 800 |
| 801 | TACTGCACACCAACAAAACACAAAGAAATTTCCCATTAATGAAAAAGAAAT | 850 |
| 306 | yTrIleLeuSerSerLeuArgAsnGlnLeuSerSerGlnLysSerValPro | 322 |

```

|||||
851 ACATCTTCATCATTAAGAATCAGCTTCTTCACAGAACTCAGTGGCG 900
323 TrypValProIleuLeuLysSerLeuProLeuTprAlaIleValAlaIaI 339
901 TGGGTACCATTTTAAATCCCTGCACATTTGGGCTATCGTAGTGCACA 950
339 sPheSerLysAntTrpThrPheTyrThrLeuLeuThrLeuLeuProThr 356
951 CTTTCTTACACACTGGACTTTTACTTATTACATTTATGCTTACTT 1000
356 yrmLtySGluIleLeuArgPheAsnValGlnGluAsnGlyPheLeuSer 372
1001 ATATGAGAGAGATCCCTAAGTTCAATGTTCAAGAAATGGGTTTATCT 1050
373 SerLeuProTyrLeuGlySerTyrLeuLysMetIleLeuSerGlyGlnAl 389
1051 TCATTTGCTTATTTAGGCTTTGGTTATGTATGATCCCTGTGTCAGAC 1100
389 aAlaSPasnLeuArgAlaLysTrpAsnPheSerThrLeuLysValArg 406
1101 TGCAGACATTTAAGCGCAAAATGGAATTTTCAACTTTATGTTGCGCA 1150
406 rglIlePheSerLeuIleGlyMetIleGlyProAlaValPheLeuValAla 422
1151 GAATTTTACGCTTATAGAAATGATGACCTGCAGATATCTGTGTAAGT 1200
423 AlaGlyPheIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrI 439
1201 GCTGGCTTCATTTGGCTGTGATTTATTTGGCGCTGTCTTCCCTAACAT 1250
439 eSerThrThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisL 456
1251 ATCAACAACACTGGAGAGCTTTTCTCTCTCTGATTTAGCATCAACCATC 1300
456 euAspIleAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThr 472
1301 TGGATATTTGCTCTTCGATCTGTATCCCTCGGCATTCACAAATACAA 1350
473 PheAlaThrIleProGlyMetValGlyProAlaIleAlaLysSerLeuTh 489
1351 TTTCCTCTATTTCCAGAAATGGTGGGCGCTCATGCTTAAATCTGCAC 1400
489 rProAspAsnThrValGlyLysTrpGlnThrValPheTyrIleAlaIaa 506
1401 CCTGATTAACACTGTGGAGAAATGCAACCGGTCTCATATTCGTCTGCTG 1450
506 lAlaIleAsnValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyLu 522
1451 CTATTTAATGTTTTGGTGGCCATTTTCTTACACTATTCCGCAAGGTGAA 1500
523 ValGlnAsnTrpAlaLeuAsnAspHisHisGlyHisArgHis 536
1501 GTCAAAAACTGGGCTCTCATGATCACCATGACACAGACAC 1542
seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AA158115
seq_documentation_block:
ID AA158115 standard: cDNA; 1488 BP.
XX
XX AA158115;
XX
XX 22-OCT-2001 (first entry)
DE Human polynucleotide SEQ ID NO 318.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia; ss.
XX

```

```

OS Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
XX
XX P-PSDB; AAM38959.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 318; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localized neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: immune system suppression,
XX CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX
XX Sequence 1488 BP; 368 A; 318 C; 323 G; 479 T; 0 other;
XX
XX
XX alignment_scores:
XX Quality: 2614.00 Length: 495
XX Ratio: 5.291 Gaps: 0
XX Percent Similarity: 99.798 Percent Identity: 99.798
XX
XX alignment_block:
XX US-09-776-865-2 x AA158115 ..
XX
XX Align seq 1/1 to: AA158115 from: 1 to: 1488
XX
XX 42 MetArgSerProValArgAspLeuAlaArgAsnAspGlyGlnGluSerThr 58
XX |||||||
XX 1 ATGAGGCTCCGCTTCAGACAGCTGCGCCGGAACGATGCGCAGAGAGAC 50
XX
XX 58 rAspArgThrProLeuLeuProGlyAlaProArgAlaGlnAlaAlaProV 75
XX |||||||
XX 51 GGACCGGACCCCTCTTCTACCGGGCGCCGACGCGGAGAACCGCTCCAG 100
XX |||||||
XX 75 aLysCysSerAlaArgTyrAsnLeuAlaIleLeuAlaPhePheGlyPhe 91
XX |||||||
XX 101 TGTGCTGCTGCTGCTTACAACTTACAAATTTTGGCTTTTGGTTTC 150
XX

```

92 PheIleValIyralaLeuArgValAsnLeuSerValAlaLeuValAspMe 108
 151 TTCTATGCTGATGCTATGCTGTAATCTGAGTGTGCTTAGTGGATAT 200
 108 tValAspSerAsnThrThrLeuGluuAspAsnArgThrSerLysAlaCysp 125
 201 GGTAGATTCAAAATACACTTAGAAGATATAGAACTTCCAGAGCGGTGC 250
 125 roGluHisSerAlaProIleLysValHisAsnGlnThrGlyLysLys 141
 251 CAGAGCATTTGCTGCCATAAAGTTTCATATCATATCAAGGGATGAAG 300
 142 TyrgIInTPaspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTy 158
 301 TACCAATGGGATGCAGAACTCAAGAGATGATTCCTCGTCTTTTATTA 350
 158 rGlyTyrlleIleThrgIInIleProGlyGlyTyrlValAlaSerLysIleG 175
 351 TGGCTACATATCACACAGATTCCTGGAGGATATGTTGCCAGCAAAATAG 400
 175 IyGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaValIleuThr 191
 401 GGGGAAAAAGCTGCTAGGATTTGGGATCCTTGGCACTGCTGCTGACC 450
 192 LeuPheThrProIleAlaAlaAspLeuGlyValGlyProLeuIleValle 208
 451 CTGTTCACCTCCCATGCTGCGAGATTAGAGTTGGACCACTCATTTACT 500
 208 uArgAlaLeuGluGlyLeuGlyGlyValIleThrPheProAlaMetHisA 225
 501 CAGAGCCTGAAGAGCTAGAGAGGGGTGTACATTTCCAGCCATGATG 550
 225 IamETTPSerSerTPAlaProProLeuGluArgSerLysLeuLeuSer 241
 551 CCATGAGTGTCTTCTGGGCTCCCTCTTAAGAAACCAACTTCTTAGC 600
 242 ILeSerTyrlaGlyAlaGlnLeuGlyThrValIleSerLeuProLeuSe 258
 601 ATTTCAATATCAGAGACAGACAGCTTGGAGACAGTAATTTCTCTCTTC 650
 258 rGlyIleIleCysTyrlMetAsnTrpThrTyrlValPheTyrlPhePheG 275
 651 TGAATATATTTGCTACTATATGAAATGGACTTATGTCCTTACTTTTGG 700
 275 IyThrIleGlyIlePheTrpPheLeuLeuTrpIleTrpLeuValSerAsp 291
 701 GTACTATATGGAATATTTGGTTCTTTTGGATCTGGATTAGTTAGTAC 750
 292 ThrProGlnLysHisLysArgIleSerHisTyrlGluLysGluTyrlIle 308
 751 ACACCAAAACACACAGAGATTTCCCATTTATGAAGAAAGAAATACATTC 800
 308 uSerSerLeuArgAsnGlnLeuSerSerGlnLysSerValProTrpValP 325
 801 TTCTATCATTTAAGAAATCAGCTTCTTCACAGAGTCCAGGCCCTGGGTAC 850
 325 roIleLeuLysSerLeuProLeuTrpAlaIleValAlaHisPheSer 341
 851 CCATTTTAAATCCCTGCCACCTTGGGCTATCTAGTTGACACATTTTCT 900
 342 TyraSnTrpThrPheTyrlThrLeuLeuThrLeuLeuProThrTyrlMet 358
 901 TACAACCTGGACTTTTATCTTATGACATTTATGGCTCTCTATATATGA 950
 358 sGluIleLeuArgPheAsnValGlnGluAsnGlyPheLeuSerSerLeuP 375
 951 GGAATCTCAAGGCTCATGTTCAAGAGAAVGGGTTTTATCTTCATGTC 1000
 375 roTyrlLeuGlySerTrpLeuCysMetIleLeuSerGlyGlnAlaAlaasp 391
 1001 CTATATTAGGCTCTGCTATGATATATCTGCTGCTCAAGCTGCTGAC 1050
 392 AsnLeuArgAlaLysTrpAsnPheSerThrLeuCysValArgArgIlePhe 408

1051 AATTTAAGGCCAAATGCAATTTTCAACTTATGCTGTGCAAAATTTT 1100
 408 eSerLeuIleGlyMetIleGlyProAlaValPheLeuValAlaIleGlyP 425
 1101 TAGCTTATAGGAATGATGATGACCTGCAGATTTCTGTGAGCTGCTGCT 1150
 425 heIleGlyCysAspTyrlSerSerLeuAlaValAlaPheLeuThrIleSerThr 441
 1151 TCATTTGGCTGTGATATTTCTTTGGCGCTGTGCTTCCATACATATTCACA 1200
 442 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuaspII 458
 1201 ACATGGGAGGCTTTGCTCTCTTCGATTTAGCATCAACCATCTGAGTAT 1250
 458 eAlaProSerTyrlAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaT 475
 1251 TGCTCTTCTGTATGCTGTGATCCTCTGGGCAATCACAATATGATTTCCA 1300
 475 hrIleProGlyMetValGlyProValIleAlaLysSerLeuThrProasp 491
 1301 CTATTCAGGAATGCTTGGGCCGTCTATTCCTAAAGTCTGACCCCTGAT 1350
 492 AsnThrValGlyGluTrpGlnThrValPheTyrlleAlaAlaIleAs 508
 1351 AACACTGTGAGAAATGCAACCGGTCTCTATATATTCCTGCTGATTTAA 1400
 508 nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnA 525
 1401 TGTTTTGGTGCATTTTCTTTACACTATTTCGCCAAAGGTGAAGTACAAA 1450
 525 snTPAlaLeuAsnAspHisHisGlyHisArgHis 536
 1451 ACTGGGCTCTCAATGATCACCATGAGACAGACAC 1485
 seq_name: /SIDSI/gcgsdata/hold-geneseq/geneseg-emb1/NA2000.DAT:AA250876
 seq_documentation_block:
 ID AA250876 standard; cDNA; 2844 BP.
 AC AA250876;
 DT 31-MAY-2000 (first entry)
 DE Sheep GBS toxin receptor (SP55) cDNA.
 KW Sheep GBS toxin receptor; group B beta-haemolytic streptococci; SP55;
 KW pathological vascularisation; cancer metastases; angiogenesis;
 KW neovascularisation; reperfusion injury; scarring; keloid;
 KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
 KW endothelial cell proliferation; antibacterial; anticancer;
 KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.
 OS Ovis sp.
 FH Key Location/Qualifiers
 FT CDS 84..1571
 FT /tag= a
 FT /product= "Sheep GBS toxin receptor"
 PN WO200005375-A1.
 PD 03-FEB-2000.
 PF 22-JUL-1999; 99WO-US16676.
 PR 22-JUL-1998; 98US-0093843.
 PA (UYVA-) UNIV VANDERBILT.
 PI Hellerqvist CG, Fu C;
 DR WPI, 2000-205377/18.


```
506 aileasnvalpheglyalalepetherleuphealalyglyluv 523
|||||
1478 TATCAATGATTTGGTCCATTTCTTCACACTATTGCCAAGGTGAAG 1527
523 aalgnasnrtpalaleuasnaspshisglvhisarghis 536
|||||
1528 TGCAGAACTGGCCATCAGTACACCAAGACAGAAAC 1568

seq_name: /sids1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT: AAD10326
seq_documentation_block:
ID AAD10326 standard; DNA; 2844 BP.
XX
AC AAD10326;
XX
DT 16-OCT-2001 (first entry)
XX
DE Sheep group B beta-haemolytic Streptococci toxin receptor (SP55) DNA.
XX
KW Sheep: group B beta-haemolytic Streptococci toxin receptor; GBS; SP55;
KW cytostatic; vulnerability; antiatherosclerotic; osteopathic; vasotropic;
KW prevention; attenuation; pathoangiogenic condition; cancer; scar;
KW wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;
KW keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;
KW vaccine; ds.
XX
OS Ovis sp.
XX
FH Key Location/Qualifiers
FT CDS 84..1571
FT /tag= a
FT /product= "Sheep GBS toxin receptor protein (SP55)"
XX
PN WO200156598-A2.
XX
PD 09-AUG-2001.
XX
PF 02-FEB-2001; 2001WO-US03662.
XX
PR 02-FEB-2000; 2000US-0179870.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
PI Hellerqvist CG;
XX
DR WPI: 2001-488844/53.
XX
DR P-PSDB: AAE06519.
XX
PT Preventing or attenuating pathoangiogenic conditions e.g. cancer,
PT chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by
PT administering group B beta-hemolytic Streptococci toxin receptor or its
PT fragment -
XX
PS Disclosure; Page 47-50; 52pp; English.
XX
XX The present sequence is a DNA encoding sheep group B beta-haemolytic
XX Streptococci (GBS) toxin receptor protein, SP55. The present invention
XX relates to a method for preventing or attenuating a patho-angiogenic
XX condition in a mammal which comprises administering to the mammal one
XX or more GBS toxin receptors or their immunogenic fragments to induce
XX CC method is useful for preventing or ameliorating pathoangiogenic
XX conditions such as cancer, scarring during wound healing, gliosis
XX during repair of nerve injury, chronic wounds, keloids, reperfusion
XX injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and
XX CC psoriasis in mammals. The proteins of the invention are also used
XX as vaccines.
XX
SQ Sequence 2844 BP; 745 A; 623 C; 594 G; 882 T; 0 other;
```

alignment_scores:

```
Quality: 2338.00 Length: 497
Ratio: 4.901 Gaps: 0
Percent Similarity: 95.976 Percent Identity: 86.117

alignment_block:
US-09-776-865-2 x AAD10326 ..
Align seg 1/1 to: AAD10326 from: 1 to: 2844

40 GLYValMetArgSerProValArgAspLeuAlaArgAsnAspGlyIugl 56
|||||
78 GGCAATCATGAAGTCCCGGTTTCGACTTAGCCCGAGCGAGCGGAGGA 127
56 userThrAspArgThrProleuLeuProGlyAlaProArgAlaIua 73
|||||
128 GGGCTCGGACGGCACACCGCTCCTGCAGCGCCGCCGGCGGAGACCCG 177
73 laProValCysSerAlaArgTyrAsnLeuAlaIleuAlaIlePhe 89
|||||
178 CTCACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 227
90 GLYPhehelValIleValAlaLeuArgValAsnLeuSerValAlaLeu 106
|||||
228 GGTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 277
106 laSpmetValAspSerAsnThrThrLeuGluAspAsnArgThrSer 123
|||||
278 GGACATGCTGATTCACAAACAACTGCCAAAGATAATAGAACCTCCAC 327
123 laCysProGluHisSerAlaProIleLysValHisAsnGlnThrGly 139
|||||
328 AGTGTGCGAGACATTCCTGCTCCATAAAGTCTTCACAAACGAGGT 377
140 LysIleGlyThrGlnThrAspAlaGluThrGlnGlyTyrIleLeuLeu 156
|||||
378 AAAAAGTACCGGGTGGATGCAAACTCAAGATGATTCGCGATCTTT 427
156 ePheTyrGlyTyrIleIleThrGlnIleProGlyIleTyrValAlaSer 173
|||||
428 TTTCTATGCTACATCATCACAAATTCCTGGAGGATATGTCGCCGCA 477
173 ysIleGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaVal 189
|||||
478 GAAGTGGGGGGAAGCTGCTAGGATTCGGAATTCCTGCTACAGCTATC 527
190 LeuThrLeuPheThrProIleAlaAlaAspLeuGlyAlaIleProleu 206
|||||
528 TTCACCTGTTCACTCCCTCCGTCGAGATTCGAGTCGAGACCCCTGT 577
206 eValLeuArgAlaLeuGluIleLeuGlyIleValIleThrProAla 223
|||||
578 TGCACTCAGGCACTAGAGGGCTAGAGAGGCTGTCACATATCCAGCA 627
223 ethAlaMetTrpSerSerTrpAlaProProleuGluArgSerLysLeu 239
|||||
628 TGCATGCCATGCTGCTTCATGGGCTCCCTCTTGAAGAACAGACCT 677
240 LeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrValIleSerLeu 256
|||||
678 CTGAGTATTCATATGAGAGGAGCAACCTGGAGCACTAGTTCCTCTCC 727
256 oluSerGlyIleIleCysTyrIleMetAsnTrpThrTyrValIlePhe 273
|||||
728 TCTTTCGGAGTAAATTCCTACATATGATTCGACTATGCTTCATAT 777
273 hebPheGlyThrIleGlyIlePheThrPheLeuLeuThrIleTyrPhe 289
|||||
778 TCTTTCGCAATTTTGAATCATCTGGTTATTTATGAGATCTCTTAGT 827
290 SerAspThrProGlnLysHisLysArgIleSerHisTyrGluLysGlu 306
|||||
828 AGTATACACCAAGAACTCACAAGACATCATCTCCCTATGAAAAAGAGTA 877
```

```
306 r1leuSerSerleuArGanGlnleuSerSerGlnlysserValProt 323
|||||
878 TATCTTTCATCATTAATAAATACAGCTCTCTTCACAGAAACAGTGGCGT 927
|||||
323 rPvalProIleuIlySerleuProIleuTPAlaIleValAlaIahis 339
|||||
928 GGAATACCTATGTCTGAAATACACTGCTTGGCTATGTGCTGCACAT 977
|||||
340 PheSerTyraNtrPrThrPheTyrThrleuNtrThrleuNtrProThrTy 356
|||||
978 TTTTCTTCAACCTGAGACTTTTATACCTTTGTTGACCTTATGCTTACTTA 1027
|||||
356 rMetlySlnIleleuArGpheAsnValGlnGluAsnGlyPheleuSers 373
|||||
1028 CATGAAAGAACTCTTAAGTTCATATATCAGAGAAATGGTTTATATCTG 1077
|||||
373 erleuProTyIleuGlySerTrPheuSylleuSerGlyGlnAla 389
|||||
1078 CAGTCCCTTATTTAGTGTGTTATGATATCCTGCTGGTCAACCT 1127
|||||
390 AlaAspAsnleuArGAlaIlyStrPAsnPhSerThrleuCySValArgAr 406
|||||
1128 GCTGACATTTTAAGGCGCAAGATGGAATTTTCACTCTGTGGGTTCCGAG 1177
|||||
406 g1lePheSerleuIleGlyMetIleGlyProAlaValPheleuValAla 423
|||||
1178 AGTTTATGACCTTATAGGATGATGAGCTGCAATATCTCTGTTGCGC 1227
|||||
423 IagIlyPheIleGlyCySAspTySerleuAlaValAlaPheleuThrIle 439
|||||
1228 CAGATTTTATAGGCTGTGATTTATTCCTGCTGTGCTTCTTAACCATTA 1277
|||||
440 SerThrThrleuGlyGlyPheCySserSerGlyPheSerIleAsnHisLe 456
|||||
1278 TCACAAACCCGAGGAGGCTTTGCTCTTCTGATTTACCAACACATCT 1327
|||||
456 uAspIleAlaProSerTyraIagIlyIleleuLeuGlyIleThraNtrPr 473
|||||
1328 GGAATTTCTCTCTGATGCTGATTTCTCTGCTGGCATCCAAATATCT 1377
|||||
473 heAlaThrIleProGlyMetValGlyProValIleAlaIlySerleuThr 489
|||||
1378 TTGCACATATCTGGAATGATGAGTGGCCCATCATTTGCCAGAGCTTAC 1427
|||||
490 ProAspAsnThrValGlyIleTrPAlnThrValPheTyrIleAlaIah 506
|||||
1428 CCTGAGAACTATTTGGAGATGCAAACTTTTCTGCATCGCTGCTGC 1477
|||||
506 alIeAsnValPheGlyAlaIlePhePheThrleuPheAlaIylselyGlu 523
|||||
1478 TATCAATGATATTTGGTGCATTTCTTCACACTATTCGCCCAAGGTGAG 1527
|||||
523 alGlnAsnTPAlaIleuAsnAspHisGlyHisARGHis 536
|||||
1528 TGCAAACTGGGCAATCAGTATCACCAAGACACAGAAAC 1568
|||||
seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-embd1/NA2000.DAT.AA250880
seq_documentation_block:
ID AA250880 standard; cDNA: 1485 BP.
XX
AC AA250880;
XX
DT 31-MAY-2000 (first entry)
XX
DE Human/Sheep consensus GBS toxin receptor cDNA sequence-1.
XX
KW Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59;
KW pathological vascularisation; cancer metastases; angiogenesis;
KW neovascularisation; reperfusion injury; scarring; keloid;
KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW endothelial cell proliferation; antibacterial; anticancer;
KW anti-angiogenic; anti-inflammatory; anti-arthritic; anti-psoriatic; ss.
```

```
XX Homo sapiens.
OS
XX Ovis sp.
FH
XX key
FT mat_peptide
FT
FT Location/Qualifiers
1..1485
/*tag_a "Human/sheep consensus GBS toxin receptor
protein sequence"
/transl_except- (pos:4..6; aa:Xaa)
/transl_except- (pos:7..9; aa:Xaa)
/transl_except- (pos:16..18; aa:Xaa)
/transl_except- (pos:22..24; aa:Xaa)
/transl_except- (pos:28..30; aa:Xaa)
/transl_except- (pos:31..33; aa:Xaa)
/transl_except- (pos:34..36; aa:Xaa)
/transl_except- (pos:46..48; aa:Xaa)
/transl_except- (pos:49..51; aa:Xaa)
/transl_except- (pos:58..60; aa:Xaa)
/transl_except- (pos:61..63; aa:Xaa)
/transl_except- (pos:64..66; aa:Xaa)
/transl_except- (pos:67..69; aa:Xaa)
/transl_except- (pos:70..72; aa:Xaa)
/transl_except- (pos:73..75; aa:Xaa)
/transl_except- (pos:79..81; aa:Xaa)
/transl_except- (pos:85..87; aa:Xaa)
/transl_except- (pos:91..93; aa:Xaa)
/transl_except- (pos:100..102; aa:Xaa)
/transl_except- (pos:124..126; aa:Xaa)
/transl_except- (pos:130..132; aa:Xaa)
/transl_except- (pos:136..138; aa:Xaa)
/transl_except- (pos:154..156; aa:Xaa)
/transl_except- (pos:157..159; aa:Xaa)
/transl_except- (pos:163..165; aa:Xaa)
/transl_except- (pos:169..171; aa:Xaa)
/transl_except- (pos:181..183; aa:Xaa)
/transl_except- (pos:187..189; aa:Xaa)
/transl_except- (pos:190..192; aa:Xaa)
/transl_except- (pos:196..198; aa:Xaa)
/transl_except- (pos:202..204; aa:Xaa)
/transl_except- (pos:211..213; aa:Xaa)
/transl_except- (pos:220..222; aa:Xaa)
/transl_except- (pos:223..225; aa:Xaa)
/transl_except- (pos:235..237; aa:Xaa)
/transl_except- (pos:241..243; aa:Xaa)
/transl_except- (pos:244..246; aa:Xaa)
/transl_except- (pos:250..252; aa:Xaa)
/transl_except- (pos:277..279; aa:Xaa)
/transl_except- (pos:280..282; aa:Xaa)
/transl_except- (pos:283..285; aa:Xaa)
/transl_except- (pos:295..297; aa:Xaa)
/transl_except- (pos:304..306; aa:Xaa)
/transl_except- (pos:337..339; aa:Xaa)
/transl_except- (pos:340..342; aa:Xaa)
/transl_except- (pos:346..348; aa:Xaa)
/transl_except- (pos:367..369; aa:Xaa)
/transl_except- (pos:394..396; aa:Xaa)
/transl_except- (pos:397..399; aa:Xaa)
/transl_except- (pos:406..408; aa:Xaa)
/transl_except- (pos:409..411; aa:Xaa)
/transl_except- (pos:412..414; aa:Xaa)
/transl_except- (pos:421..423; aa:Xaa)
/transl_except- (pos:430..432; aa:Xaa)
/transl_except- (pos:433..435; aa:Xaa)
/transl_except- (pos:436..438; aa:Xaa)
/transl_except- (pos:442..444; aa:Xaa)
/transl_except- (pos:445..447; aa:Xaa)
/transl_except- (pos:463..465; aa:Xaa)
/transl_except- (pos:475..477; aa:Xaa)
/transl_except- (pos:481..483; aa:Xaa)
/transl_except- (pos:487..489; aa:Xaa)
/transl_except- (pos:490..492; aa:Xaa)
/transl_except- (pos:493..495; aa:Xaa)
```



```
FT /transl_except- (pos:496..498; aa:Xaa)
FT /transl_except- (pos:502..504; aa:Xaa)
FT /transl_except- (pos:514..516; aa:Xaa)
FT /transl_except- (pos:529..531; aa:Xaa)
FT /transl_except- (pos:535..537; aa:Xaa)
FT /transl_except- (pos:562..564; aa:Xaa)
FT /transl_except- (pos:589..591; aa:Xaa)
FT /transl_except- (pos:595..597; aa:Xaa)
FT /transl_except- (pos:598..600; aa:Xaa)
FT /transl_except- (pos:604..606; aa:Xaa)
FT /transl_except- (pos:619..621; aa:Xaa)
FT /transl_except- (pos:634..636; aa:Xaa)
FT /transl_except- (pos:655..657; aa:Xaa)
FT /transl_except- (pos:691..693; aa:Xaa)
FT /transl_except- (pos:694..696; aa:Xaa)
FT /transl_except- (pos:700..702; aa:Xaa)
FT /transl_except- (pos:703..705; aa:Xaa)
FT /transl_except- (pos:706..708; aa:Xaa)
FT /transl_except- (pos:712..714; aa:Xaa)
FT /transl_except- (pos:715..717; aa:Xaa)
FT /transl_except- (pos:724..726; aa:Xaa)
FT /transl_except- (pos:727..729; aa:Xaa)
FT /transl_except- (pos:736..738; aa:Xaa)
FT /transl_except- (pos:748..750; aa:Xaa)
FT /transl_except- (pos:757..759; aa:Xaa)
FT /transl_except- (pos:760..762; aa:Xaa)
FT /transl_except- (pos:769..771; aa:Xaa)
FT /transl_except- (pos:772..774; aa:Xaa)
FT /transl_except- (pos:775..777; aa:Xaa)
FT /transl_except- (pos:778..780; aa:Xaa)
FT /transl_except- (pos:790..792; aa:Xaa)
FT /transl_except- (pos:793..795; aa:Xaa)
FT /transl_except- (pos:811..813; aa:Xaa)
FT /transl_except- (pos:820..822; aa:Xaa)
FT /transl_except- (pos:847..849; aa:Xaa)
FT /transl_except- (pos:850..852; aa:Xaa)
FT /transl_except- (pos:853..855; aa:Xaa)
FT /transl_except- (pos:856..858; aa:Xaa)
FT /transl_except- (pos:862..864; aa:Xaa)
FT /transl_except- (pos:880..882; aa:Xaa)
FT /transl_except- (pos:883..885; aa:Xaa)
FT /transl_except- (pos:892..894; aa:Xaa)
FT /transl_except- (pos:928..924; aa:Xaa)
FT /transl_except- (pos:928..930; aa:Xaa)
FT /transl_except- (pos:943..945; aa:Xaa)
FT /transl_except- (pos:952..954; aa:Xaa)
FT /transl_except- (pos:955..957; aa:Xaa)
FT /transl_except- (pos:970..972; aa:Xaa)
FT /transl_except- (pos:994..996; aa:Xaa)
FT /transl_except- (pos:997..999; aa:Xaa)
FT /transl_except- (pos:1009..1011; aa:Xaa)
FT /transl_except- (pos:1012..1014; aa:Xaa)
FT /transl_except- (pos:1033..1035; aa:Xaa)
FT /transl_except- (pos:1063..1065; aa:Xaa)
FT /transl_except- (pos:1081..1083; aa:Xaa)
FT /transl_except- (pos:1084..1086; aa:Xaa)
FT /transl_except- (pos:1090..1092; aa:Xaa)
FT /transl_except- (pos:1096..1098; aa:Xaa)
FT /transl_except- (pos:1111..1113; aa:Xaa)
FT /transl_except- (pos:1126..1128; aa:Xaa)
FT /transl_except- (pos:1129..1131; aa:Xaa)
FT /transl_except- (pos:1138..1140; aa:Xaa)
FT /transl_except- (pos:1141..1143; aa:Xaa)
FT /transl_except- (pos:1144..1146; aa:Xaa)
FT /transl_except- (pos:1147..1149; aa:Xaa)
FT /transl_except- (pos:1150..1152; aa:Xaa)
FT /transl_except- (pos:1153..1155; aa:Xaa)
FT /transl_except- (pos:1168..1170; aa:Xaa)
FT /transl_except- (pos:1174..1176; aa:Xaa)
FT /transl_except- (pos:1180..1182; aa:Xaa)
FT /transl_except- (pos:1189..1191; aa:Xaa)
FT /transl_except- (pos:1201..1203; aa:Xaa)
FT /transl_except- (pos:1246..1248; aa:Xaa)
```

```
FT /transl_except- (pos:1270..1272; aa:Xaa)
FT /transl_except- (pos:1291..1293; aa:Xaa)
FT /transl_except- (pos:1306..1308; aa:Xaa)
FT /transl_except- (pos:1315..1317; aa:Xaa)
FT /transl_except- (pos:1324..1326; aa:Xaa)
FT /transl_except- (pos:1330..1332; aa:Xaa)
FT /transl_except- (pos:1333..1335; aa:Xaa)
FT /transl_except- (pos:1339..1341; aa:Xaa)
FT /transl_except- (pos:1348..1350; aa:Xaa)
FT /transl_except- (pos:1357..1359; aa:Xaa)
FT /transl_except- (pos:1372..1374; aa:Xaa)
FT /transl_except- (pos:1375..1377; aa:Xaa)
FT /transl_except- (pos:1381..1383; aa:Xaa)
FT /transl_except- (pos:1384..1386; aa:Xaa)
FT /transl_except- (pos:1396..1398; aa:Xaa)
FT /transl_except- (pos:1402..1404; aa:Xaa)
FT /transl_except- (pos:1420..1422; aa:Xaa)
FT /transl_except- (pos:1444..1446; aa:Xaa)
FT /transl_except- (pos:1456..1458; aa:Xaa)
FT /transl_except- (pos:1459..1461; aa:Xaa)
FT /transl_except- (pos:1462..1464; aa:Xaa)
FT /transl_except- (pos:1471..1473; aa:Xaa)
FT /transl_except- (pos:1483..1485; aa:Xaa)
FT /note= "Xaa = unknown"
```

MO200005375-A1.

03-FEB-2000.

22-JUL-1999; 99WO-US16676.

22-JUL-1998; 98US-0093843.

(UYVA-) UNIV VANDERBILT.

Hellergvist CG, Fu C;

WPI; 2000-205377/18.

P-PSDB; AAU45090.

alignment_scores: Quality: 2217.00 Length: 494

Ratio: 5.073 Gaps: 0

Percent Similarity: 88.462 Percent Identity: 86.437

alignment_block:

US-09-776-865-2 x AAZ50880 ..

Align seg 1/1 to: AAZ50880 from: 1 to: 1485

```
42 MetArgSerProValArgAspLeuAlaArgAsnAspIylGluSerTh 58
1 ATGARGTCYCCGGTTYSRGACYTRGCCSGARCGAYGCGAGGRCGWC 50
58 RAspArgThProLeuLeuProGlyAlaProArgAlaGlaAlaAlaPro 75
51 GGACCGCACRCKCTYCTRCMGSGCGCCRCRGCGGCGAASCGCTCAG 100
92 PheLeuValAlaArgValAsnLeuSerValAlaLeuAlaPhe 91
101 TRGCTGCTGCGCGTTACACACTTGTGCKCTTTTGGCTTC 150
92 PheLeuValAlaArgValAsnLeuSerValAlaLeuAlaPhe 108
151 TTCCTTSTRTATTCATTACGCTGAATCTGAGTGCTTCRTAGGAYAT 200
108 tValAspSerAsnThrThrLeuGluAspAsnArgThSerLysAlaGly 125
201 GGTGATTCAAAYACACTYKMAAGATATATAGACTCCWASMGCTIS 250
125 roGluHisSerAlaProIleLysValHisHisasnGlnThrGlyLys 141
```

| | | |
|------|---|------|
| 251 | CAGACATCTCTGCTCCCATTAAGAAAGTTCTWTCATAAATCAAAAGGGTAAARAG | 300 |
| 142 | TyrGlnTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePhe | 158 |
| 301 | TACCRRTRGGAGTGCAGAAATCAAGATGGATTCTCGWTCYTTTYYTA | 350 |
| 158 | rgLYrYrIleIleThrGlnIleProGlyIyYrYrValAlaSerIysIleG | 175 |
| 351 | TGGCTACATCATCCACACAAATTCCTGGAGGAATATGTGTGCACACAAARWG | 400 |
| 175 | IyGlyIysMetLeuLeuGlyIlePheGlyIleLeuGlyThrAlaValLeuThr | 191 |
| 401 | GGGGGCAARTRGTYGCTAGATATTTGGGATCTTGTGATACMGCTCTCTACCC | 450 |
| 192 | LeuPheThrProIleAlaIleAspLeuGlyValGlyProLeuIleValLe | 208 |
| 451 | CTGTTCACATCCGCTTGTCTGCAGATTMTMGAGTYGGASGCTYRTGYACT | 500 |
| 208 | uArGAlaLeuGlyIleGlyIleGlyIleGlyValThrPheProAlaMetHisA | 225 |
| 501 | CARGCACATGAAAGRCCTAGAGAGGGTGTACAAWTCCAGCCATGACATG | 550 |
| 225 | lMetIrrSerSerTrpAlaProProLeuGluArgSerIysIleuLeuSer | 241 |
| 551 | CCATGGGTCTCTWTTGGGTCCCCCTCTTGAAAGAACACARCTTCTKAGY | 600 |
| 242 | lIleSerTyralaGlyAlaGlnLeuGlyIleThrValIleSerLeuProLeuSe | 258 |
| 601 | ATTTCRRATGACAGACACACARCTGGGACAGATRTTCTCTTCCTCTTC | 650 |
| 258 | rgLYIleIleCysTyTYrMetAsnTrpThrTyValPheTyPhePheG | 275 |
| 651 | TGGARPAATTTGCTACTAATATGAATGGACTATGTCTTCATYTTTGTG | 700 |
| 275 | LYrHrIleGlyIlePheTrpPheLeuLeuTrpIleTrpLeuValSerasp | 291 |
| 701 | GIATYRTTGGAAATWMTYTGTTTTTTTTRTGGATCTGTTAGTTAGTAY | 750 |
| 292 | ThrProGlnIleHisIysAspArgIleSerHisIstYrGlnIysGluTyrIleLe | 308 |
| 751 | ACACCAASAAMWCACAGASASATWYCMYRTATGAAAAGATATATCT | 800 |
| 308 | uSerSerLeuArgAsnGlnLeuSerSerGlnIysSerValProTrpValP | 325 |
| 801 | TYTCATCATTAARAANAACACACTYCTTCTACAGAAACAGTCCGCTGRTAC | 850 |
| 325 | rolIleLeuIysSerLeuProLeuTrpAlaIleValValAlaHisPheSer | 341 |
| 851 | CYATKTYRAAATCMCGCCACTTGGGCTATYATGATGTGCACATYTTTCT | 900 |
| 342 | TYrAsnTrpThrPheTyTrhLeuLeuThrIleLeuLeuProThrTyMetIy | 358 |
| 901 | TACAACTGGACTTTTATATCTTTTGTGACMTTATTTGCTACTTAYATGAA | 950 |
| 358 | sgLIuIleLeuArgPheAsnValGlnGluAsnGlyPheLeuSerLeuP | 375 |
| 951 | GGARTRCTAAGGTCAATRTTCAAGAGAAATGGTTTTATCTCKAKATSC | 1000 |
| 375 | roTYrIleuIySerTrpLeuCysMetIleLeuSerGlyGlnAlaAlaasp | 391 |
| 1001 | CTTATTTAGAGTSTTGGTATATGATGATCTCTCGGTCAGACCTCTAC | 1050 |
| 392 | AsnLeuArgAlaIystrPaspAsnPheSerThrIleCysValAlaGargIlePh | 408 |
| 1051 | AAATTTAAGGGCACAARAATTTTTCACCTYTRTGKGTTCGMAAGARTTTT | 1100 |
| 408 | eSerIleuIleGlyMetIleGlyProAlaValPheLeuValAlaAlaGlyP | 425 |
| 1101 | TACCTTTATAGGATATGATTGGACATGRCRRATTTCTGHWGCGWCGMT | 1150 |
| 425 | heIleGlyCysAspTYrSerLeuAlaValAlaPheLeuThrIleSerThr | 441 |
| 1151 | TYTWTMGCTGTGATATYTTCTYTGCGYGTGTGCTTCTCTACATATATACACA | 1200 |

| | | |
|---|--|------|
| 442 | ThrlenglylphcysSerSerglyPheSerlleasnhtlsleuapil | 458 |
| | | |
| 1201 | ACMCTGGAGGCTTTTGCCTTCCTCGATTAGCATCAACATCCTGGAYAT | 1250 |
| 458 | ealProSetTyralaglIyleuleauglyilethAsnThrPheAlat | 475 |
| | | |
| 1251 | TGCTCCCTTGTAIGCTGTATCTCCCGGGCATCACAAATRCMTTGGCA | 1300 |
| 475 | hrlleProglYmetValgIyProValllleaalySerrleuthrProasp | 491 |
| | | |
| 1301 | CFAATTCMGGAATGRITGGGCCCCRCATCTGCYAANAAGCTKACCCCTGAK | 1350 |
| 492 | AsnThrValalgIutlrpglnInthrValPheTyrrllealaalaileas | 508 |
| | | |
| 1351 | AACACTTTTGAGATGGCAACACGTGKTTRATVCTCTGCTATYAA | 1400 |
| 508 | nValPheGlYAlaIllePhePhehrleauPheAlaValgIylgluValGlna | 525 |
| | | |
| 1401 | TGWTITTGGTGCCATTTTCTTYACACTATTTCGCCAAGAAGTGAACTRCAAA | 1450 |
| 525 | sntTPALaleuasnASPHisHisGlyHIsarg | 535 |
| | | |
| 1451 | ACTGGGCGYMCARTGATCACACAGACACAGA | 1482 |
| seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ50881 | | |
| seq_documentation_block: | | |
| ID | AAZ50881 standard; cDNA; 1485 BP. | |
| XX | | |
| AC | AAZ50881; | |
| DT | | |
| DE | 31-MAY-2000 (first entry) | |
| XX | | |
| Human/sheep consensus GBS toxin receptor cDNA sequence-2. | | |
| Human GBS toxin receptor: group B beta-haemolytic streptococci; HP59; | | |
| pathological vasculatisation: cancer metastases; angiogenesis; | | |
| KW neovascularisation; reperfusion injury; scarring; keloid; | | |
| KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury; | | |
| KW endothelial cell proliferation; antibacterial; anticancer; | | |
| anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss | | |
| XX Homo sapiens. | | |
| OS Ovis sp. | | |
| XX | | |
| Key Location/Qualifiers | | |
| FH mat_peptide | 1..1485 | |
| FT *tag= a | /product= "Human/sheep consensus GBS toxin receptor | |
| FT protein sequence" | | |
| FT /transl_except= (pos:4..6; aa:Xaa) | | |
| FT /transl_except= (pos:7..9; aa:Xaa) | | |
| FT /transl_except= (pos:16..18; aa:Xaa) | | |
| FT /transl_except= (pos:22..24; aa:Xaa) | | |
| FT /transl_except= (pos:28..30; aa:Xaa) | | |
| FT /transl_except= (pos:31..33; aa:Xaa) | | |
| FT /transl_except= (pos:34..36; aa:Xaa) | | |
| FT /transl_except= (pos:46..48; aa:Xaa) | | |
| FT /transl_except= (pos:49..51; aa:Xaa) | | |
| FT /transl_except= (pos:58..60; aa:Xaa) | | |
| FT /transl_except= (pos:61..63; aa:Xaa) | | |
| FT /transl_except= (pos:64..66; aa:Xaa) | | |
| FT /transl_except= (pos:67..69; aa:Xaa) | | |
| FT /transl_except= (pos:70..72; aa:Xaa) | | |
| FT /transl_except= (pos:73..75; aa:Xaa) | | |
| FT /transl_except= (pos:79..81; aa:Xaa) | | |
| FT /transl_except= (pos:85..87; aa:Xaa) | | |
| FT /transl_except= (pos:91..93; aa:Xaa) | | |
| FT /transl_except= (pos:100..102; aa:Xaa) | | |
| FT /transl_except= (pos:124..126; aa:Xaa) | | |
| FT /transl_except= (pos:130..132; aa:Xaa) | | |
| FT /transl_except= (pos:136..138; aa:Xaa) | | |

FT /transl_except- (pos:154..156; aa:Xaa)
FT /transl_except- (pos:157..159; aa:Xaa)
FT /transl_except- (pos:163..165; aa:Xaa)
FT /transl_except- (pos:169..171; aa:Xaa)
FT /transl_except- (pos:181..183; aa:Xaa)
FT /transl_except- (pos:187..189; aa:Xaa)
FT /transl_except- (pos:190..192; aa:Xaa)
FT /transl_except- (pos:196..199; aa:Xaa)
FT /transl_except- (pos:202..204; aa:Xaa)
FT /transl_except- (pos:211..213; aa:Xaa)
FT /transl_except- (pos:220..222; aa:Xaa)
FT /transl_except- (pos:223..225; aa:Xaa)
FT /transl_except- (pos:235..237; aa:Xaa)
FT /transl_except- (pos:241..243; aa:Xaa)
FT /transl_except- (pos:244..246; aa:Xaa)
FT /transl_except- (pos:250..253; aa:Xaa)
FT /transl_except- (pos:277..279; aa:Xaa)
FT /transl_except- (pos:280..282; aa:Xaa)
FT /transl_except- (pos:283..285; aa:Xaa)
FT /transl_except- (pos:295..297; aa:Xaa)
FT /transl_except- (pos:304..306; aa:Xaa)
FT /transl_except- (pos:337..339; aa:Xaa)
FT /transl_except- (pos:340..342; aa:Xaa)
FT /transl_except- (pos:346..348; aa:Xaa)
FT /transl_except- (pos:367..369; aa:Xaa)
FT /transl_except- (pos:394..396; aa:Xaa)
FT /transl_except- (pos:397..399; aa:Xaa)
FT /transl_except- (pos:406..408; aa:Xaa)
FT /transl_except- (pos:409..411; aa:Xaa)
FT /transl_except- (pos:412..414; aa:Xaa)
FT /transl_except- (pos:421..423; aa:Xaa)
FT /transl_except- (pos:430..432; aa:Xaa)
FT /transl_except- (pos:433..435; aa:Xaa)
FT /transl_except- (pos:436..438; aa:Xaa)
FT /transl_except- (pos:442..444; aa:Xaa)
FT /transl_except- (pos:445..447; aa:Xaa)
FT /transl_except- (pos:463..465; aa:Xaa)
FT /transl_except- (pos:475..477; aa:Xaa)
FT /transl_except- (pos:481..483; aa:Xaa)
FT /transl_except- (pos:487..489; aa:Xaa)
FT /transl_except- (pos:490..492; aa:Xaa)
FT /transl_except- (pos:493..495; aa:Xaa)
FT /transl_except- (pos:496..498; aa:Xaa)
FT /transl_except- (pos:502..504; aa:Xaa)
FT /transl_except- (pos:514..516; aa:Xaa)
FT /transl_except- (pos:529..531; aa:Xaa)
FT /transl_except- (pos:535..537; aa:Xaa)
FT /transl_except- (pos:562..564; aa:Xaa)
FT /transl_except- (pos:589..591; aa:Xaa)
FT /transl_except- (pos:595..597; aa:Xaa)
FT /transl_except- (pos:598..600; aa:Xaa)
FT /transl_except- (pos:604..606; aa:Xaa)
FT /transl_except- (pos:619..621; aa:Xaa)
FT /transl_except- (pos:634..636; aa:Xaa)
FT /transl_except- (pos:655..657; aa:Xaa)
FT /transl_except- (pos:691..693; aa:Xaa)
FT /transl_except- (pos:694..696; aa:Xaa)
FT /transl_except- (pos:700..702; aa:Xaa)
FT /transl_except- (pos:703..705; aa:Xaa)
FT /transl_except- (pos:706..708; aa:Xaa)
FT /transl_except- (pos:712..714; aa:Xaa)
FT /transl_except- (pos:715..717; aa:Xaa)
FT /transl_except- (pos:724..726; aa:Xaa)
FT /transl_except- (pos:727..729; aa:Xaa)
FT /transl_except- (pos:736..738; aa:Xaa)
FT /transl_except- (pos:748..750; aa:Xaa)
FT /transl_except- (pos:757..759; aa:Xaa)
FT /transl_except- (pos:760..762; aa:Xaa)
FT /transl_except- (pos:769..771; aa:Xaa)
FT /transl_except- (pos:772..774; aa:Xaa)
FT /transl_except- (pos:775..777; aa:Xaa)
FT /transl_except- (pos:778..780; aa:Xaa)
FT /transl_except- (pos:790..792; aa:Xaa)

FT /transl_except- (pos:793..795; aa:Xaa)
FT /transl_except- (pos:811..813; aa:Xaa)
FT /transl_except- (pos:820..822; aa:Xaa)
FT /transl_except- (pos:847..849; aa:Xaa)
FT /transl_except- (pos:850..852; aa:Xaa)
FT /transl_except- (pos:853..855; aa:Xaa)
FT /transl_except- (pos:856..858; aa:Xaa)
FT /transl_except- (pos:862..864; aa:Xaa)
FT /transl_except- (pos:880..882; aa:Xaa)
FT /transl_except- (pos:883..885; aa:Xaa)
FT /transl_except- (pos:892..894; aa:Xaa)
FT /transl_except- (pos:922..924; aa:Xaa)
FT /transl_except- (pos:928..930; aa:Xaa)
FT /transl_except- (pos:943..945; aa:Xaa)
FT /transl_except- (pos:952..954; aa:Xaa)
FT /transl_except- (pos:955..957; aa:Xaa)
FT /transl_except- (pos:970..972; aa:Xaa)
FT /transl_except- (pos:994..996; aa:Xaa)
FT /transl_except- (pos:997..999; aa:Xaa)
FT /transl_except- (pos:1009..1011; aa:Xaa)
FT /transl_except- (pos:1012..1014; aa:Xaa)
FT /transl_except- (pos:1033..1035; aa:Xaa)
FT /transl_except- (pos:1063..1065; aa:Xaa)
FT /transl_except- (pos:1081..1083; aa:Xaa)
FT /transl_except- (pos:1084..1086; aa:Xaa)
FT /transl_except- (pos:1090..1092; aa:Xaa)
FT /transl_except- (pos:1096..1098; aa:Xaa)
FT /transl_except- (pos:1111..1113; aa:Xaa)
FT /transl_except- (pos:1126..1128; aa:Xaa)
FT /transl_except- (pos:1129..1131; aa:Xaa)
FT /transl_except- (pos:1138..1140; aa:Xaa)
FT /transl_except- (pos:1141..1143; aa:Xaa)
FT /transl_except- (pos:1144..1146; aa:Xaa)
FT /transl_except- (pos:1147..1149; aa:Xaa)
FT /transl_except- (pos:1150..1152; aa:Xaa)
FT /transl_except- (pos:1153..1155; aa:Xaa)
FT /transl_except- (pos:1168..1170; aa:Xaa)
FT /transl_except- (pos:1174..1176; aa:Xaa)
FT /transl_except- (pos:1180..1182; aa:Xaa)
FT /transl_except- (pos:1189..1191; aa:Xaa)
FT /transl_except- (pos:1201..1203; aa:Xaa)
FT /transl_except- (pos:1246..1248; aa:Xaa)
FT /transl_except- (pos:1270..1272; aa:Xaa)
FT /transl_except- (pos:1291..1293; aa:Xaa)
FT /transl_except- (pos:1306..1308; aa:Xaa)
FT /transl_except- (pos:1315..1317; aa:Xaa)
FT /transl_except- (pos:1324..1326; aa:Xaa)
FT /transl_except- (pos:1330..1332; aa:Xaa)
FT /transl_except- (pos:1333..1335; aa:Xaa)
FT /transl_except- (pos:1339..1341; aa:Xaa)
FT /transl_except- (pos:1348..1350; aa:Xaa)
FT /transl_except- (pos:1357..1359; aa:Xaa)
FT /transl_except- (pos:1372..1374; aa:Xaa)
FT /transl_except- (pos:1375..1377; aa:Xaa)
FT /transl_except- (pos:1381..1383; aa:Xaa)
FT /transl_except- (pos:1384..1386; aa:Xaa)
FT /transl_except- (pos:1396..1398; aa:Xaa)
FT /transl_except- (pos:1402..1404; aa:Xaa)
FT /transl_except- (pos:1420..1422; aa:Xaa)
FT /transl_except- (pos:1444..1446; aa:Xaa)
FT /transl_except- (pos:1456..1458; aa:Xaa)
FT /transl_except- (pos:1459..1461; aa:Xaa)
FT /transl_except- (pos:1462..1464; aa:Xaa)
FT /transl_except- (pos:1471..1473; aa:Xaa)
FT /transl_except- (pos:1483..1485; aa:Xaa)
/note="Xaa = unknown"

WO200005375-A1.

03-FEB-2000.

22-JUL-1999; 99WO-US16676.

XX 20-NOV-2001 (first entry)
XX Human sodium dependent phosphate co-transfer protein 35 cDNA.
DE Human sodium dependent phosphate co-transfer protein 35;
XX hypophosphaturia; hypercalcaemia; hypophosphataemic rickets;
KW nephritis; gene therapy; ss.
XX Homo sapiens.
XX CN1298882-A.
XX 13-JUN-2001.
XX PD
XX 06-DEC-1999; 99CN-0124217.
XX PR 06-DEC-1999; 99CN-0124217.
XX PA (SHEN-) SHENGYUAN GENE DEV CO LTD SHANGHAI.
XX PI Mao Y, Xie Y;
XX DR MPI: 2001-503367/56.
XX DR P-PSDB: AAG65238.
XX PT Human Na-dependent phosphate cotransporter 35 and its coding sequence -
XX PS Claim 6; Page 19-20(Disclosure); 28pp; Chinese.
XX CC The present invention provides the protein and coding sequences of human
XX CC sodium dependent phosphate co-transfer protein 35. The sequences can be
XX CC used in the treatment of hypophosphaturia, hypercalcaemia,
XX CC hypophosphataemic rickets and nephritis. The present sequence is the
XX CC coding sequence of the invention.
SQ Sequence 2670 BP; 706 A; 550 C; 552 G; 862 T; 0 other;

alignment_scores:
Quality: 1716.00 Length: 321
Ratio: 5.346 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-776-865-2 x AAH79234 ..

Align seg 1/1 to: AAH79234 from: 1 to: 2670

216 GIJGLYVALThrPheProAlaMetHisAlaMetTTPSerSerTIPAlaPr 232
|||||
3 GAGGGGTTCATTTCCAGCCATGCATGCCATGCGTCTTGGGCTCC 52

232 oProLeuGluArgSerLysLeuSerLysSerTyrAlaGlyAlaGlnL 249
|||||
53 CCCTCTGGAAGAAGAACCTTCTAGCATTCATATGACAGACAGACAC 102

249 euGLYThrValIleSerLeuProLeuSerGlyIleIleCysTyrTrpMet 265
|||||
103 TTGGGACAGTAATTTCTCTCTCTTCTGGAATTAATTTGCTACTATATG 152

266 AsnTrpThrValPheTyrPhePheGlyTrpIleGlyIlePheTrpH 282
|||||
153 AATTGGACTTATGCTCTTCTACTTTTGTGCTACTATGGAATATTTGGTT 202

282 eLeuLeuTrpIleTyrPheValSerAspThrProGlnLysHisLysArgI 299
|||||
203 TCTTTTGGGATCTGGTAGTAGTGACACACACAAAAACACAGAGAGA 252

299 lSerHisTyrGlnLysGlnTyrIleLeuSerSerLeuArgAsnGlnLeu 315
|||||
253 TTTCCTATTATGAAAAAGAAATACATTTCTTCATCATTAAGAATACGCTT 302

316 SerSerGlnLysSerValProTyrValProIleLeuLysSerLeuPro 332
|||||
303 TCTTCACAGAACTCAGTCCGCTGGGTACCCATTTTAAATCCCTGCCACT 352

332 uTPAlaIleValAlaHisPheSerTyrAsnTrpThrPheTyrThrL 349
|||||
353 TTGGGCTATCGTAGTTCACACTTCTTCAACAACGTGACTTTTATACCT 402

349 euLeuThrLeuLeuProThrTyrMetLysGlnIleLeuArgPheAsnVal 365
|||||
403 TATTGACATTAATGGCTACTTATATGAGAGAGATCCTAAGTTCAATGTT 452

366 GlnGluAsnGlyPheLeuSerSerLeuProTyrLeuGlySerTrpLeu 382
|||||
453 CAAGAGAAATGGGTTTATCTTCATTCATTCCTTATTAAGCTCTTGTTAG 502

382 sMetIleLeuSerGlyAlaAlaAspAsnLeuArgAlaLysTrpAsn 399
|||||
503 TATGATCCTGTCTGTGCAAGCTGCTGACAATTTAAGGCCAAATGGAAT 552

399 hSerThrLeuLysValArgArgIlePheSerLeuIleGlyMetIleGly 415
|||||
553 TTTCACCTTATGCTTCCAGAAATTTTACCTTATAGGAATGATTTGA 602

416 ProAlaValPheLeuValAlaAlaGlyPheIleGlyCysAspTyrSer 432
|||||
603 CTTGCAGTATTCCTGTGTGCTGCTGCTGCTTATTCATTCATTCATTC 652

432 uAlaValAlaPheLeuThrIleSerThrThrLeuGlnGlyPheCysSer 449
|||||
653 GCGCGTGTCTTCTTACTATATCAACACACTGGAGAGCTTTGCTCTT 702

449 euGLYPheSerIleAsnHisLeuAspIleAlaProSerTyrAlaGlyIle 465
|||||
703 CTGGATTTAGCATCAACCACTGGATTTGCTCTGCTGCTGCTGCTGCTAT 752

466 LeuLeuGlyIleThrAsnThrPheAlaThrIleProGlyMetValGlyPr 482
|||||
753 CTCCTGGGATCAACAATACATTTGCCACTATTCACGAAATGTTGGGCC 802

482 oValIleAlaLysSerLeuThrProAspAsnThrValGlyIleTyrGlnT 499
|||||
803 CGTCATTTGCTAAAGCTGACCCCTGATTAACACTGTTGGAGATGGCAAA 852

499 hrValPheTyrIleAlaAlaIleAsnValPheGlyAlaIlePhePhe 515
|||||
853 CCGTGTCTTATATGCTGCTGCTATTAATGTTTGTGCTCATTTCTTCT 902

516 ThrLeuPheAlaLysGlyGluValGlnAsnTrpAlaLeuAsnAspHisH 532
|||||
903 ACACATATTCGCCAAAGGTAAAGTACAAAACCTGGCTCTCAATGATCACC 952

532 sGLYHisArgHis 536
|||||
953 TGGACACAGACAC 965

seq_name: /SIS1/gcdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI59901
seq_documentation_block:
ID AAI59901 standard; cDNA; 929 BP.
XX
XX AAI59901;
AC
XX
DT 22-OCT-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 3890.
DE
XX Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.

XX OS Homo sapiens.
XX PN WO200153312-A1.
XX PD 26-JUL-2001.
XX PF 26-DEC-2000; 2000MO-US34263.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 29-NOV-2000; 2000US-0727344.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX DR WPI: 2001-442253/47.
XX DR P-PSDB: AAM40745.
XX PT Novel nucleic acids and polypeptides, useful for treating disorders
XX PT such as central nervous system injuries -
XX PS Claim 1: SEQ ID NO 3890; 10078bp; English.
XX CC The invention relates to human nucleic acids (AA157798-AA161369) and
XX CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful
XX CC in gene therapy. A composition containing a polypeptide or polynucleotide
XX CC of the invention may be used to treat diseases of the peripheral nervous
XX CC system, such as peripheral nervous injuries, peripheral neuropathy and
XX CC localised neuropathies and central nervous system diseases, such as
XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX CC utilisation of the activities such as: Immune system suppression,
XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX CC assays for receptor activity, arthritis and inflammation, leukaemias and
XX CC C.N.S disorders.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification.
XX SQ Sequence 929 BP; 214 A; 224 C; 223 G; 268 T; 0 other;

alignment_scores:
Quality: 1614.00 Length: 304
Ratio: 5.309 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-776-865-2 x AA159901 ..

Align seg 1/1 to: AA159901 from: 1 to: 929

8 ProProArgProValGlnProAlaArgProGlyGlyPheGlyLeuSerG1 24
17 CCGGCCCGCCGGGCCAGCGAGCTCGGGCGGCTTCGGGCTGTGGG 66
24 YArgArgSerLeuLeuGlyGlnValAlaSerThrProAlaHisValGly 41
67 CCGGCGCTCCCTTCTCTGCGAGGTGGCGAGTACACCTGCTCAGTAGCG 116
41 AlMetArgSerProValArgAspLeuAlaArgAsnAspGlyGluGluSer 57
117 TCATGAGTCTCGGGTTCGAGACCTGGCCGGAACGATGGCGAGAGAGC 166

58 ThrAspArgThrProLeuLeuProGlyValAlaProArgAlaGluAlaApr 74
167 ACGGACCGCAGCGCTCTTCTACCGGGGCCCCACGCGGCCGAGCGCTCC 216
74 OvaLysCysSerAlaArgTyrAsnLeuAlaIleLeuAlaPhePheGlyP 91
217 AGTGTGCTGCTCTGCTGTTACCACTTAGCAATTTTGGCTTTTGTGTT 266
91 hePheIleValTyrAlaLeuArgValAsnLeuSerValAlaLeuValAsp 107
267 TCTTCATTTGTTATGATGCAATGCTGATCGAATCGAATGCTGGCTAGTGA 316
108 MetValAspSerAsnThrThrLeuGluAspAsnArgThrSerLysAla 124
317 ATGCTACATTCATCAATACACTTTAGAAAGATATAGAACTTCCAAAGCGCTG 366
124 sProGluHisSerAlaProIleLysValHisHisAsnGlnThrGlyLysL 141
367 TCCAGACCATTTCTGCTCCATMAAAGTTCAATCAATCAAAACGGGTAGA 416
141 YstYrGlnTTPAspAlaGluThrGlnGlyThrIleLeuGlySerPhePhe 157
417 AGTACCAATGGGATGAGAAATCAAGATGATTCGCGTCTCTTTT 466
158 TyrGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerLysI 174
467 TATGGCTACATCATCACACAGATTCCTGGAGGATTTGTCGACGACAAAT 516
174 eGlyGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaValLeu 191
517 AGGGGGGMAAATGCTGCTAGATTGGGATCTTGGGACGCTGTCTCTCA 566
191 hrLeuPheThrProIleAlaAlaAspLeuGlyValGlyProLeuIleVal 207
567 CCTGTTCACTCCCATTTGCTGCGAGATTTAGGAGTTGGACACATCATTTG 616
208 LeuArgAlaLeuGlyGlyLeuGlyGlyValThrPheProAlaMetH 224
617 CTGAGACACTAGAGAGAGTAGAGAGGAGGTTACATTTCCACGACATGCA 666
224 sAlaMetThrSerSerThrProAlaProLeuGluArgSerLysLeuLeu 241
667 TGGCATTTGCTCTTCTGGGCTCCCTCTTGAAGAAGCAAACTTCTTA 716
241 erLleSerTyrAlaGlyAlaGlnLeuGlyThrValIleSerLeuProLeu 257
717 GCATTTGATATGACAGACAGCTTGGACAGTAATTTCTTCTCTCTT 766
258 SerGlyIleIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPheP 274
767 TCTGGAATATATTGCTACATATGAAATGAGCTTATGCTTCTTACTTTT 816
274 eGlyThrIleGlyIlePheThrPheLeuLeuTrpIleTrpLeuValSera 291
817 TGGTACTATTGGAATATTTTGGTTTCTTTTGTGATCTGCTGTAGTAGT 866
291 sPThrProGlnLysHisLysArgIleSerHisTyrGlnLysGluTyrIle 307
867 ACACACCACAAAAACACAGAGAAATTTCCCATTTATGAAAAAGAAATACAT 916
308 LeuSerSerLeu 311
917 CTTTTCATCATTTA 928

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseg-emb1/NA2001A.DAT:AAK93901
seq_documentation_block:
ID AAK93901 standard; cDNA; 853 BP.
XX AAK93901;
XX 06-NOV-2001 (first entry)

```

XX Human cDNA clone representative sequence, SEQ ID NO: 2361.
DE
XX
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
KM
XX
OS Homo sapiens.
XX
PN EP130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
PI use in genetic manipulation -
XX
PS Example 11: SEQ ID NO 2361; 1380pp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence was used as the
CC representative sequence from a human clone which was used in
CC homology searches to identify the clone.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 853 BP; 198 A; 200 C; 213 G; 239 T; 3 other;

alignment_scores:
      Quality: 1417.00      Length: 284
      Ratio: 5.079          Gaps: 0
Percent Similarity: 98.239  Percent Identity: 97.183

alignment_block:
US-09-776-865-2 x AAK93901 ..

Align seg 1/1 to: AAK93901 from: 1 to: 853

15 AAlarPProGlyIyPheGlyLeuSerGlyAlrArgSerLeuLeuGyl 31
   |||||||
2  GCTGGGCCCGGGGCTTCGGGCTTCGGGCCGCGCTCCCTCTCTGCA 51
   |||||||
31 nValAlaSerThrProAlaHisValGlyValMetArgSerProValArg 48
   |||||||
52 GGTGGCAGATACACCTCTCAGTAGCGTCATGAGGTCTCCGGTTCGAG 101
   |||||||
48 spLeuAlaArgAsnAspGlyGluGluSerThrAspArgThrProLeuLeu 64
   |||||||
102 ACCTGGCCCGGAACGATGGCGAGAGACGACGCGCCAGCCCTCTCTTA 151
   |||||||
65 ProGlyAlaProArgAlaGluAlaAlaProValCysCysSerAlaArgTy 81
   |||||||
152 CCGGGGCCCGCCACGCGCGGAAGCGCTCCAGTGTGCTGCTGCTCTTA 201
   |||||||
81 rAsnLeuAlaIleLeuAlaIlePhePheGlyPhePheIleValTyraLeu 98

```

```

|||||
202 CAACCTAGCAATTTGGCTTTTGGTTCTTCATGTGTATGCATTCAC 251
   |||||||
98 rGValAsnLeuSerValAlaLeuValAspMetValAspSerAsnThrThr 114
   |||||||
252 GTGTGAATCTGAGTGTGCTTCGTTAGTGAATGTGTAATTCAAATACAC 301
   |||||||
115 LeuGluAspAsnArgThrSerLysAlaCysProGluHisSerAlaProI 131
   |||||||
302 TTGAAGATATATAGACTTCACAGCGGTGTCACAGCATTTGTGCTCCAT 351
   |||||||
131 elysValHisHisAsnGlnThrGlyLysTyrGlnTrpAspAlaGluT 148
   |||||||
352 AAAAGTTCATCATATATCAAAAGGGTAGAAGTACCAATGGATGCAGAAA 401
   |||||||
148 hrgGlnGlyTrpIleLeuGlySerPhePheGlyGlyTyrIleIleThrGln 164
   |||||||
402 CTCAGGATGATGATTCGCTCTCTTTTATGCTCATCATCATCACAG 451
   |||||||
165 IleProGlyGlyTyrValAlaSerLysIleGlyGlyLysMetLeuLeuG 181
   |||||||
452 ATTCTTGAGGATATGTTCGACAGCAAAATAGGGGGGAAATGCTGTAGG 501
   |||||||
181 yPheGlyIleLeuGlyThrAlaValAlaLeuThrLeuPheThrProIleAla 198
   |||||||
502 ATTTGGGATCCTTGGCAGCTGCTGCTCCACCCCTGTCACCTGCCATTCCT 551
   |||||||
198 laAspLeuGlyValGlyProLeuIleValLeuArgAlaLeuGlyLeu 214
   |||||||
552 CAGATTTAGAGAGTTGGACCACTCATTTGACTCAGAGCAGTAGAAGGACTA 601
   |||||||
215 GlyGluGlyValThrPheProAlaMetHisAlaMetTrpSerSerTrpAl 231
   |||||||
602 GGAAGAGGCTGTACATTTCCAGCCATCATGCCATGTGCTGCTTTTGGCC 651
   |||||||
231 aProProLeuGluArgSerLysLeuLeuSerLysSerTyrAlaGlyLac 248
   |||||||
652 TCCCCCTTTGAAAGAGCAAACTTTAGCATTTTCATATGACAGAGACAC 701
   |||||||
248 lInLeuGlyThrValIleSerLeuProLeuSerGlyIleIleCysTyrTyr 264
   |||||||
702 AGCTTGGACAGATATTTCTCTCTCTCTGGAATATTTGCTACTAT 751
   |||||||
265 MetAsnTrpThrTyrValPheTyrPhePheGlyThrIleGlyIlePheTr 281
   |||||||
752 ATGAATTTGACTTATGCTCTTCTACTTTTGTGNACTATGTGAAAATTTG 801
   |||||||
281 pPheLeuLeuTrpIleTrpLeuValSerAspThrProGluLysHisLysA 298
   |||||||
802 GGTCTTTTNGGATCTNGGTAGT.AGTGACACACACCAAAAACCCGAGA 850
   |||||||
298 rg 298
   ||
851 GA 852

seq_name: /SIDSI/gcgdata/hold-geneseg/geneseqn-emb1/NA2001A.DAT:AAK92364
seq_documentation_block:
ID AAK92364 Standard; cDNA; 838 BP.
XX
XX AAK92364;
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human cDNA 5'-end sequence, SEQ ID NO: 824.
DE
XX
XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
KM
XX
OS Homo sapiens.
XX
PN EP130094-A2.
PD 05-SEP-2001.

```

XX 07-JUL-2000; 2000EP-0114089.
 PF 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX (HELI-) HELIX RES INST.
 PA Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX WPI; 2001-524255/58.
 DR 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 PS Claim 2; SEQ ID NO 824; 1380pp + sequence listing; English.
 XX The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is the nucleotide
 CC sequence of the 5'-end of a cDNA provided in the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX
 SQ Sequence 838 BP; 189 A; 196 C; 210 G; 237 T; 6 other;
 alignment_scores:
 Quality: 1339.50 Length: 281
 Ratio: 4.998 Gaps: 2
 Percent Similarity: 95.374 Percent Identity: 94.306
 alignment_block:
 US-09-776-865-2 x AAK92364 ..
 Align seg 1/1 to: AAK92364 from: 1 to: 838

131 elysValHisHisnngInhrGlyLysTyrcInlrpAspAlaGut 148
 |||||
 352 AAAGTTCATCATATATAAAGCGGTAAAGACGTACCAATGGATGCGAATA 401
 148 hrGInGlyTrpIleLeuGlySerPhePheTyTYrIleIleThrGln 164
 |||||
 402 CTCAGGATGATGATCTCGTTCCTTTTATAGCTATCATCATACACAG 451
 165 IleProGlyGlyTYrValAlaSerTyIleGlyGlyIleMetLeuLeu 181
 |||||
 452 ATTCCTGGAGATATATGTTCCAGCAAAATAGGGGGGAAATGCTCTG 501
 181 yPheGlyIleLeuGlyThrAlaValLeuThrLeuPheThrProIleAla 198
 |||||
 502 ATTTGGATCCCTGGACCTGCTGCTCCACCCCTGCTACCTCCATTGCTG 551
 198 laAspLeuGlyValGlyProLeuIleValIleuArgAlaLeuGlyLeu 214
 |||||
 552 CAGATTTAGGAGTTGGACCACTCATTTGACTCAGACACTAGAGACTA 601
 215 GlyGlyGlyValThrPheProAlaMetHisAlaMetTrpSerSerTrpAl 231
 |||||
 602 NGAGAGGCTGTACATTTCCAGCCATGCATGCATGTGTCTTCTGGGC 651
 231 aProProLeuGlyuArgSerLysLeuLeuSerLysSerTyraIleGlyAlaG 248
 |||||
 652 TCCTCCCTTGAAAGAACCAACTTTCATGATTTCAATATCAGAGAC 701
 248 InLeuGlyThrValIleSerLeuProLeuSerGlyIleIleCysTyTYr 264
 |||||
 702 ANCTTGGGACAGTAATTTCTTCTTCTTCTGGAATATATATGCTACTAT 750
 265 MetAsnTrpPheThrValPheTyPhePheGlyThrIleGlyIlePheTr 281
 |||||
 751 ATGATTTGGACTTATGCTTCTACTTTTGGCATATATGGAATATTTGG 800
 281 PheLeuLeuTrp.IleTrpLeuValSerAspPheProGln 294
 |||||
 801 GTTC...TTTGGGAGACTGNTAANTANTGACACACCCCAA 838
 seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn_emb1/NA2001B.DAT:ABU03769
 seq_documentation_block:
 ID ABL03769 standard; cDNA; 1939 BP.
 XX
 AC ABL03769;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 5789.
 XX
 KW Drosophila: developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 PD 27-SEP-2001.
 XX
 PD 23-MAR-2001; 2001WO-0509231.
 XX
 PF 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EW;
 DR WPI; 2001-656860/75.
 DR P-PDB; ABB59666.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1: SEQ ID NO 5789; 21bp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL161/6-ABL30511), expressed DNA
CC sequences (AB101840-ABL161/5) and the encoded proteins
(AB57737-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pt_sequences.

SQ Sequence 1939 BP; 410 A; 558 C; 552 G; 419 T; 0 other;

alignment_scores:

| Quality: | 996.50 | Length: | 523 |
|---------------------|--------|-------------------|--------|
| Ratio: | 2.815 | Gaps: | 10 |
| Percent Similarity: | 67.686 | Percent Identity: | 39.771 |

alignment_block:

US-09-776-865-2 x ABL03769 ..

Align seg 1/1 to: ABL03769 from: 1 to: 1939

```
51 ArgAsnAspGlyGluGluSerThrAspArg..... 60
   |||||  |||  ::  ::|||
350 CGCTCCGACGACGAGCGAGACGAGGAGCGTTTGTCTCCGCGCA 399
   |||||  |||||  |||||  |||||  |||||  |||||
61 .ThrProLeuProGlyAlaProAlaGluAlaAlaProValCys. 76
   |||||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
400 ACGACCGCTATCCGCTCCAGCGCGTCCGAGAAACCATGCGTGTG 449
   |||||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
77 .CysSerAlaArgTyrAsnLeuAlaIleLeuAlaPheGlyPhePhe 92
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
450 GCCCAAGAGCGGTCACATTTTCGATTCATGGCTTCCTGGGATTCGC 499
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
93 IleValTyrAlaLeuArgValAsnLeuSerValAlaLeuValAspMet 109
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
500 GTGGTCTACGCGATGCGGTCAATCTGCGTGGCCATGTGGCCATGCT 549
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
109 Lasp.....SerAsnThrThrLeuGluAspAsnArgT 120
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
550 GAACCAAGCGCAATTCGACAGCAACATCATCGGTGATGATGACGACA 599
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
120 hSerTyrAlaCysProGluHisSerAlaProIleLeuValHisHisAsn 136
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
600 CG.....TGTCGA...CTACCGGACCA.....CATCACAAT 628
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
137 .....GlnThrGlyLysTyrGlnTrpAspAlaG 147
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
629 GGTAGCGATCCCAATCCGACAGAGGAGCGAGTTGTGTGGGACGAGCG 678
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
147 uThrGlnGlyTrpIleLeuGlySerPhePheTyrGlyTyrIleLeuThr 164
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
679 CAGCGAGAGTGTGTGCTGCGCAGATTCCTTAAGCTATGCTTAACCC 728
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
164 InIleProGlyGlyTyrValAlaSerLysIleGlyGlyLysMetLeu 180
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
729 AAGTGGCCGCGGAGCGATGCGGAGCTGATGTGGAGAGATCTAC 778
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
181 GlyPheGlyIleLeuGlyThrAlaValLeuThrLeuPheThrProIle 197
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
779 GCGTATGAGTGTGATACACGCGGCTTAACTTAATCACTCATGTCG 828
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
197 aaIleAspLeuGlyValGlyProLeuIleValLeuArgAlaLeuGlu 214
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
829 TGCCCACTGGATCTGCCGCTGTGTGTCGTGTCGCATCTCTGGAGGAA 878
```

```
214 euGIgLIuGIValThrPheProAlaMetHisAlaMetTrpSerSerTrp 230
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
879 TGGCGAGGGCGCTACCTATCCAGCTATGCACGCCATGCTGCCACTGG 928
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
231 AlaProLeuGluArgSerLysLeuLeuSerLysSerTyrAlaGlyAl 247
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
929 ATACCGCGCGCTGGAGAGAACAGTTCGCCGACATCTCTATGCGGGCTC 978
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
247 agIleuGlyThrValIleSerLeuProLeuSerGlyIleLeuCys... 262
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
979 CAATATCGAACAGTCAATTCATCCGCGTGGCGGATGCTGTGCTCGC 1028
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
263 .....TyrTyrMetAsnTrpThrTyrValPheTyrPhePheGlyThr 277
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1029 TGGACTTCTCGGGGCTGGCTGGCCGCTTCCTTCAATCTTCGAGCTGTG 1078
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
278 GlyIlePheThrPheLeuLeuTrpIleTrpLeuValSerAspThrPro 294
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1079 GGCATTCGTGTGATCATCGCATGATGTATTTGGTGTACGACCAAGCCTAG 1128
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
294 nLysHisLysArgIleSerHisTyrGluLysGluTyrIleLeuSerSer 311
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
1129 CGATCATCCAGGATCTCAGAAATCGAGCGAGATATATCGAAGAGATC 1178
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
311 euArgAsnGln..... 314
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1179 TACAGTTCAGAGGCTAATATCAGATCTACGCGAGCGCGAGAGAGAG 1228
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
315 .....LeuSerSerGlnLysSerValPr 322
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1229 GAGGAGCAGATGAAAGTGTGTCGCGGCGCGCGGAGGAGACCGATACC 1278
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
322 cTrpValProIleLeuLysSerLeuProLeuTrpAlaIleValAlaHis 339
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1279 CTGGTCATCGCTGCTACATCCGATCCTGTGTGGCCATCTGTGTGGCG 1328
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
339 LspSerSerTyrAsnTrpThrPheTyrThrLeuLeuThrLeuProThr 355
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1329 AATGCGGCCAGGAGTGGCGCTTCTACACGACGACTAACCGAGCTGCCACC 1378
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
356 TyrMetLysGluIleLeuArgPheAsnValGlnGluAsnGlyPheLeu 372
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1379 TACATGAGCAACATCCACATTTGACATCCAGCGAATGCTGCTGCAC 1428
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
372 rSerLeuProTyrLeuGlySerTrpLeuCysMetIleLeuSerGlyGln 389
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1429 TGGGTGCTCGATCTAACCCTCGGTTCGTGGGCAATTCCTGCTCCGCC 1478
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
389 LaAlaAspAsnLeuArgAlaLysTrpAsnPheSerThrLeuCysValArg 405
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1479 TGGCGATTTGGATGCTGACGACGACGCTACATTCGCTGGAATCTGAT 1528
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
406 ArgIlePheSerLeuIleGlyMetIleGlyProAlaValAlaPheLeu 422
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1529 AAGTTTGGAAACAGCGTGGCTCCGTGTCATCACTGAGGCTGATGG 1578
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
422 aaIleGlyPheIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThr 439
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1579 CATCATCTATGTGGCGTGGATGGGTGTGTGACCTTTATGTGTGGCG 1628
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
439 LserThrThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHis 455
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1629 GCGTGGGCTCTTCGCGGGCGCGCTCATCTGCAACCAAGATGATCAC 1678
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
456 LeuAspIleAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsn 472
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1679 ATAGCGCTAGTCCAGCATATGACGACCATGATATGACATCAACATTC 1728
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
472 rPheAlaThrIleProGlyMetValGlyPro...ValIleAlaLysSer 488
   |||||  |||||  |||||  |||||  |||||  |||||  |||||
1729 GCGGCAAAATATCTGTGGCTTCGTGCTCGTATGTCATGATGATATCA 1778
```


OM of: US-09-776-865-2 to: EST:* out_format: pfs

Date: Jul 15, 2002 2:32 PM

About: Results were produced by the Gencore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+2.p2n.model -DEV=xlh
-O=/cgn2.1/USPRO.spool/US09776865/runat.15072002.082207.27382/app-query.fasta.1.1159
-DB=EST -QMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blsnum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=200 -THR MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09776865@cgn2.1.3468
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-776-865-2

Database: EST*

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 2434.030000

score_list:

| Sequence | Strd Orig | ZScore | EScore | Len | Documentation |
|------------------|-----------|---------|----------|------|----------------------------------|
| gb_est1:AL550137 | 1707.00 | 2746.90 | 9.3e-144 | 985 | AL550137 LTI_NFL006.P1 |
| gb_est2:BI907284 | 1326.00 | 1986.56 | 2.1e-101 | 754 | BI907284 60306385881 NIH_MGC_11 |
| gb_est2:BI661062 | 1115.00 | 1789.83 | 1.9e-90 | 801 | BI661062 60330436281 NIH_MGC_11 |
| gb_est2:BI868761 | 1062.00 | 1705.79 | 9.1e-86 | 676 | BI868761 6014431271 NIH_MGC_65 |
| gb_est2:BI860521 | 1024.50 | 1644.09 | 2.5e-82 | 736 | BI860521 6033867871 NIH_MGC_87 |
| gb_est2:BF676817 | 991.00 | 1591.57 | 2.1e-79 | 626 | BF676817 60208438081 NIH_MGC_83 |
| gb_est2:BI697765 | 991.00 | 1590.18 | 2.0e-79 | 711 | BI697765 60208438081 NIH_MGC_83 |
| gb_est2:BF539146 | 930.50 | 1489.64 | 1.0e-73 | 885 | BF539146 6020500991 NCI_CGAP_5 |
| gb_est2:BF14137 | 909.00 | 1454.69 | 8.8e-72 | 904 | BF14137 60176083881 NCI_CGAP_5 |
| gb_hic:BC023331 | 909.00 | 1448.95 | 1.8e-71 | 1531 | BC023331 Mus musculus, clone 1 |
| gb_est2:BG541099 | 900.00 | 1440.24 | 5.6e-71 | 833 | BG541099 60257026581 NIH_MGC_77 |
| gb_est1:BB613552 | 899.00 | 1441.55 | 4.8e-71 | 665 | BB613552 BB613552 RIKEN full-16 |
| gb_est1:BB537525 | 863.50 | 1384.63 | 7.1e-68 | 645 | BB537525 BB537525 RIKEN full-16 |
| gb_est1:AL597124 | 862.00 | 1385.00 | 6.7e-68 | 499 | AL597124 DKR2P313H1911_r1 313 |
| gb_est1:AA833297 | 802.00 | 1287.36 | 1.8e-62 | 518 | AA833297 ud05009_r1 Soares_NHNP |
| gb_est2:BE869819 | 793.00 | 1269.50 | 1.8e-61 | 700 | BE869819 6014465181 NIH_MGC_65 |
| gb_est2:BG866603 | 782.00 | 1246.75 | 3.4e-60 | 1100 | BG866603 6027855771 NCI_CGAP_1 |
| gb_est2:BI817031 | 777.50 | 1245.48 | 4.0e-60 | 633 | BI817031 UMN07804 Canine Brain |
| gb_est1:BB610013 | 772.50 | 1243.10 | 5.4e-60 | 630 | BB610013 RIKEN full-16 |
| gb_est1:BB664731 | 772.50 | 1237.44 | 1.1e-59 | 629 | BB664731 RIKEN full-16 |
| gb_est2:BI851890 | 711.50 | 1138.10 | 3.8e-54 | 658 | BI851890 6033780471 NCI_CGAP_1 |
| gb_est2:BM179717 | 707.50 | 1132.86 | 7.4e-54 | 587 | BM179717 da111810.y1 NICHD XGC |
| gb_est2:BU060500 | 682.00 | 1089.30 | 2.0e-51 | 721 | BU060500 BU060500 NIBB Mochi1 |
| gb_est1:AA258513 | 681.00 | 1093.26 | 1.2e-51 | 432 | AA258513 z759001_r1 Soares_NHNP |
| gb_est2:BF563945 | 662.00 | 1062.55 | 6.1e-50 | 429 | BF563945 UT-R-C4-akw-c-03-0-UT |
| gb_est2:BU031727 | 639.00 | 1021.06 | 1.3e-47 | 632 | BU031727 NIBB Mochi1 |
| gb_est2:BI445533 | 539.00 | 858.78 | 1.4e-38 | 647 | BI445533 dae81109.y3 NICHD XGC |
| gb_est2:BF459662 | 532.50 | 845.76 | 7.3e-38 | 811 | BF459662 AT14282.Spiral AT Dros |
| gb_est2:BE375421 | 530.00 | 848.25 | 5.3e-38 | 445 | BE375421 60123048581 NCI_CGAP_1 |
| gb_est2:BI63685 | 520.50 | 834.95 | 2.9e-37 | 367 | BI63685 yf55601_r1 Soares_fetal |
| gb_est2:BI873013 | 517.00 | 819.99 | 2.0e-36 | 866 | BI873013 60339819381 NIH_MGC_94 |
| gb_est2:BF123545 | 503.50 | 799.30 | 2.8e-35 | 772 | BF123545 6017588681 NCI_CGAP_1 |
| gb_est2:BE892556 | 499.50 | 792.99 | 6.4e-35 | 760 | BE892556 6014333581 NIH_MGC_72 |
| gb_est2:BI13154 | 494.00 | 794.19 | 5.5e-35 | 301 | BI13154 yx5306.0_r1 Soares_melan |
| gb_est2:BI734022 | 484.00 | 768.79 | 1.4e-33 | 753 | BI734022 6033555781 NIH_MGC_94 |
| gb_gss:CMS0069V | 482.00 | 757.24 | 6.2e-33 | 1075 | AL063966 Drosophila melanogast |
| gb_est2:BI753164 | 476.50 | 757.24 | 6.2e-33 | 661 | BI753164 60302028281 NIH_MGC_11 |
| gb_gss:CMS0352N | 472.00 | 746.08 | 2.6e-32 | 943 | AL223388 Tetradon nigrofasciatis |
| gb_est2:BF45471 | 461.50 | 732.64 | 1.5e-31 | 679 | BF45471 AT04410.Spiral AT Dros |
| gb_est2:BU106525 | 439.50 | 697.10 | 1.4e-29 | 672 | BU106525 BU106525 unpublished |

gb_est2:BG123873 + 438.50 694.13 2.0e-29 761 | BG123873 EST469519 tomato sh
gb_est2:BU032567 + 428.00 680.45 1.2e-28 560 | BU032567 NIBB Mochi
gb_est2:BG296175 + 428.00 676.80 1.9e-28 783 | BG296175 6023940061 NIH_MGC
gb_est2:BU148966 - 427.50 676.36 2.0e-28 757 | BU148966 BU148966 unpublished
gb_est1:AM531337 - 427.00 684.89 6.7e-29 321 | AM531337 UT-R-C4-akw-c-03-0-

seq_name: gb_est1:AL550137

seq_documentation block:

LOCUS AL550137 985 bp mRNA linear EST 16-FEB-2001
DEFINITION AL550137 LTI_NFL006.P12 Homo sapiens cDNA clone CS001040YK17 5
prime, mRNA sequence.

```

89  eglYpHePheIleValTyraIaLeuArgValAsnLeuSerValAlaLeu 106
|||||
203  TGGTTTCTTCATGTGTGATGCATTACGTGATCTGATCTGGCTGAG 252
|||||
106  aLAspMeValAspSerAsnThrThrLeuGluAspAsnArgThrSerIys 122
|||||
253  TGGATATGTGTAGATTCAAAATACAACTTTAGAAAGATAAATAGAACTTCCAG 302
|||||
123  AlaCysProGluHisSerAlaProIleLeuValHisHisAsnGlnThrGlu 139
|||||
303  GCGTCTCCAGAGCATTTCTGCTCCCAATAAAGTTCACTCAATACAAACGGG 352
|||||
139  yLysLysTyrgLnTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerP 156
|||||
353  TAAGAAGTACCAATGGATGCAGAAACTCAAGAGATTCGCGGTCTCT 402
|||||
156  hePheTyrgLyTyrrIleIleThrGlnIleProGlyGlyTyrrValAlaSer 172
|||||
403  TTTTATTAGCTACATCATCACACAGATTCTCGAGATATGTTGCCAGC 452
|||||
173  LysIleGlyGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaVal 189
|||||
453  AAATA.GGGGGGAAATGCTGCTAGATTGGGATCTTGCGACTGCTGT 501
|||||
189  lleuthrLeuPheThrProIleAlaAlaAspLeuGlyValGlyProLeuI 206
|||||
502  CCTGACCCCTGTTCACTCCCATTTGGCGAGATTTGGAGATTGGACACATCA 551
|||||
206  leValleuArgAlaLeuGluGlyLeuGlyGlyGlyValThrPheProAla 222
|||||
552  TTGACTAGAGACACTAGAAAGACTAGAGAGGGGTTTACATTTCCACCC 601
|||||
223  MethIAsaMetTrpSerSerTrpAlaProProLeuGluArgSerLysLe 239
|||||
602  ATGCATGCCATGTGCTCTTCTTGGGCTCCCTCTGTGAAGAAGCAACT 651
|||||
239  uLeuSerIleSerTyraIaGlyAlaGlnLeuGlyThrValIleSerLeuP 256
|||||
652  TCTTAGCATTTCAATATGCAGAGACAGACTTGGGACAGTAATTTCTCTTC 701
|||||
256  roLeuSerGlyIleIleLeuSerTyrrMetAsnTrpThrTyrrValPheTy 272
|||||
702  CTCCTTCTGGAAATATTTGCTACTATAGAAATGGACTTATGCTCTTAC 751
|||||
273  PhePheGlyThrIleGlyIlePheThrPheLeuLeuTrpIleTrpLeuVa 289
|||||
752  TTTTGTGCTACTATTTGAAATTTTGGTTTCTTTTGTGATCTGCTTACT 801
|||||
289  lSerAspThrProGlnLysHisLysArgIleSerHisTyrgLnuLysGlu 306
|||||
802  TAGTGACACACACAAAAACACAAGAGAAATTTCCCATATGAAAGGAAAT 851
|||||
306  yrrIleLeuSerSerLeuArgAsnGlnLeuSerSerGlnLysSerValPro 322
|||||
852  ACATCTTTCATTCATTAAGAAATCAGCTTCTTCCACAGAACTCAGTGGCG 901
|||||
323  TrpValProIleLeuLysSerLeuProLeuTrpAlaIleValAlaHis 339
|||||
902  TGGGTACCCATTTTAAATCCCTGCCACTTGGGCTATCGTAGTTGCACA 951
|||||
339  sPheSerTyraSnTrpThrPheTyrrThrLeuLeu 350
|||||
952  CTTTTCTTACAACTGRCRTTTTATACTTATTTG 985
|||||
seq_name: gb_est2:BI907284
seq_documentation_block:
LOCUS      BI907284                754 bp      mRNA      linear      EST 16-OCT-2001
DEFINITION 603063858F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5212766 5',
            mRNA sequence.
ACCESSION  BI907284
VERSION    BI907284.1  GI:16170094

```

```

KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 754)
            NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS   National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE     Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgabs@email.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LRAM1534 row: d column: 15
            High quality sequence stop: 750.
FEATURES   Location/Qualifiers
            source
            1..754
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:5212766"
                /clone_lib="NIH_MGC_118"
                /tissue_type="leukocyte"
                /lab_host="DH10B"
                /note="Vector: PCMV-SpOrf6; Site_1: NotI; Site_2: EcoRV
                (destroyed); RNA source: Leukocytes from anonymous pool of
                non-activated adult donors. Library is oligo-dT primed
                and directionally cloned (EcoRV site is destroyed upon
                cloning). Average insert size 1.7 kb, insert size range
                1.2-3.3 kb. Library is normalized and enriched for
                full-length clones and was constructed by C. Gruber
                (Invitrogen). Research genetics tracking code 027. Note:
                this is a NIH-MGC Library."
BASE COUNT 171 a 182 c 194 g 207 t
ORIGIN
alignment_scores:
    Quality: 1236.00      Length: 251
    Ratio: 4.944          Gaps: 1
    Percent Similarity: 99.602      Percent Identity: 99.203
alignment_block:
US-09-776-865-2 x BI907284 ..
Align seg 1/1 to: BI907284 from: 1 to: 754
18  GYGlyPheGlyLeuSerGlyArgArgSerLeuLeuGlyValAlaIase 34
|||||
6  GGGGGCTTGGGGCTGTGGGGCGGGCGCTCTTCTGTGCCAGGTGGGAG 55
|||||
34  rThrProAlaHisValGlyValMetArgSerProValArgAspLeuAla 51
|||||
56  TACACCTGCTCAGCTAGAGCGTCATGAGGTCTCCGGTTCCAGAACTGGGCC 105
|||||
51  rGAsnAspGlyGluGluSerThrAspArgThrProLeuLeuProGlyValA 67
|||||
106  GGAACGATGGGAGAGAGCAGCGACCG .ACGCTCTTCTTACCCGGGGGCC 154
|||||
68  ProArgIaGluAlaAlaProValCysCysSerAlaArgTyraSnLeuAl 84
|||||
155  CCACGGGCCGAGAGCCGCTCCAGTGTGCTGCTGCTGCTTACCAACTTAGC 204
|||||
84  alleLeuAlaPhePheGlyPhePheIleValTyraIaLeuArgValAsnL 101
|||||
205  AATTTGGCCCTTTTGTGTTCTTCATTTGATGATGATACGTTGATATC 254
|||||
101  euSerValAlaLeuValAspMetValAspSerAsnThrThrLeuGluAsp 117
|||||

```

```

255 TGAGTGTTCGCTAGTGGATATGTTAGATTCAAAATACAACTTTAGAGAT 304
118 AsnArgThrSerLysAlaCysProGluHisSerAlaProIleLysValH1 134
|||||
305 AATGAACTTCCAGGCGGTGCCAGACATTCTCTCCATTAAGTTCA 354
134 SHASnGlnThrGlyLysLysTyrGlnTrpAspAlaGluThrGlnLys 151
|||||
355 TCATTAATCAAAACGGGTAAAGAGTACCAATGGGATGCAAGAACTCAAGAT 404
151 rPileLeuGlySerPhePheTyrGlyTyrIleIleThrGlnIleProGly 167
|||||
405 GGATTCTCGTTCCTTTTATGCTACATCATCACACAGATTCCCTGGA 454
168 GLTYrValAlaSerLysIleGlyLysMetLeuGlyPheGlyI1 184
|||||
455 GGAATCTTCCAGCAAAATAGGGGGAATAATGCTGCTAGCATTTGGGAT 504
184 eLeuGlyThrAlaValLeuThrLeuPheThrProIleAlaAlaAspLeuG 201
|||||
505 CCTGGCAGCTGCTGCTCCTCACTGTTCACTCCATGCTGACAGATTAG 554
201 LysAlaGlyProLeuIleValLeuArgAlaLeuGluGlyLeuGly 217
|||||
555 GAGTTGGACCACTCATTTGTAAGTACAGACACTAGAGAGAGAGAGG 604
218 ValThrPheProAlaMetHisAlaMetTrpSerSerTrpAlaProProle 234
|||||
605 GTTACATTTCCAGCCATGCATGCCATGTGGCTTCTTGGGCTCCCTCT 654
234 uGluArgSerLysLeuLeuSerLysTyrAlaGlyAlaGlnLeuGly 251
|||||
655 TGAAGAAGCAAACTCTTAGCATTTATGACAGACAGAGCTTGGA 704
251 hValIle.SerLeuProLeuSerGlyIleLeuGlyTyrTyrMetAsnTr 267
|||||
705 CAGTAATTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 753
267 P 267
754 G 754

```

seq_name: gb_est2:BI661062

seq_documentation_block: 801 bp mRNA linear EST 12-SEP-2001
 LOCUS BI661062
 DEFINITION 603304362F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5350046 5',
 mRNA sequence.

ACCESSION BI661062
 VERSION BI661062.1 GI:15575298
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-r@mail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furch
 Ph.D.

CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM1889 row: 1 column: 15
 High quality sequence stop: 799.

FEATURES
 source 1..801
 Location/Qualifiers

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5350046"
/clone_1ib="NIH_CGAP_Mam4"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert 2.5 kb. Library constructed by Life Technologies, catalog # 12018-016. Investigators providing samples: Lothar Hennighausen/Priscilla Furch, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP library."
BASE COUNT 173 a 200 c 220 g 208 t
ORIGIN

```

alignment_scores:
 Quality: 1115.00 Length: 257
 Ratio: 4.705 Gaps: 4
 Percent Similarity: 92.218 Percent Identity: 83.268

alignment_block:
 US-09-776-865-2 x BI661062 ..

Align seg 1/1 to: BI661062 from: 1 to: 801

```

40 GlyValMetArgSerProValArgAspLeuAlaArgAsnAspGlyGlu 56
|||||
22 GGCCTTATGAGGCCCGCTTCGGGGTCCCGCGGAAACGCGATGAGGA 71
56 UserThrAspArgThrProLeuLeuProGlyAlaProArgIleGluAla 73
|||||
72 GAGCTCGGACAGACCCCGCTCTGCGGCGCGCGGACGACGAGAGGG 121
73 IapProValCysSerAlaArgTyrAsnLeuAlaIleLeuAlaPhe 89
|||||
122 CTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 171
172 GGTTCCTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 221
106 IAspMetValAspSerAsnThrThrLeuGluAspAsnArgThrSerLys 123
|||||
222 GGACATGTGATTCATCAATCAACTGACTGATTAATGAACGTTAGAG 271
123 IacysProGluHisSerAlaProIleLysValHisSHASnGlnThrGly 139
|||||
272 AGTGTGGGAAACATTCGCCCCCTAAAGTTCACCATCACACAGGT 321
140 LysLysTyrGlnTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerP 156
|||||
322 AAAAAGTACAAGTGGATGAGAACTCAAGGCGGATTCGCGCTCTTT 371
156 ePheTyrGlyTyrIleThrGlnIleProGlyLysTyrValAlaSerL 173
|||||
372 TTTTACGGCTACATCGTCAACCATTCGCCGGGGGTGATTCACCGCA 421
173 ySILeGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaVal 189
|||||
422 GGTGCGAGGGAAGCTGCTGCTGGCGCTGGGCACTTACGACCTCGCTC 471
190 LeuThrLeuPheThrProIleAlaAlaAspLeuGlyValGlyProLeu 206
|||||
472 TTCACCTGTTACACCGCTGCGCGGACAGACTTAGGCGTGGAGACTCT 521
206 eValLeuArgAlaLeuGluGlyLeuGlyValAlaThrPheProAla 223
|||||
522 TGTCCTTAGAGCGCTGGAAGAGCTGGGAGAGGGGTTACGTTCTCACCT 571
223 eHisAlaMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeu 239

```

```

|||||
572 TGCAGCAGATGCTCTCCCTGGGCTCCCTCTGGAAGAAGCAGCTT 621
240 LeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrValIleSerLeuPr 256
|||||
622 CTTACCAATTCCTATGCGGAGACAGCTTGAGAGATGATCTCACTTCC 671
256 OLeuSerGlyIleIleGlyTyrTyrMetAsnTyrThrTyrValPheTyr. 272
672 TCTTCCGGAATATATGCTATATGACTGAGCTTACGCTCTCTATTC 721
273 PhePheGlyThrIleGlyIlePheThrPheLeu.LeuTriPleTyrLeuV 289
|||||
722 TTT...GGTATAGTTGGAATGTCTGG...TTATTTATGATGTGTGATAG 765
289 aLeSerAspThrProGln 294
|||||
766 TCAGTGATACACACAGA 782
seq_name: gb_est2:BE867611

```

```

seq_documentation_block: 676 bp mRNA linear EST 20-OCT-2000
LOCUS BE867611 601443127F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3847279 5',
DEFINITION mRNA sequence.
ACCESSION BE867611
VERSION BE867611.1 GI:10316387
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 676)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
plate: L1AM9561 row: e column: 08
High quality sequence stop: 676.
FEATURES
Location/Qualifiers
source
1..676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3847279"
/tissue_type="adrenocarcinoma"
/lab_host="DH10B (Phage-resistant)"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site: 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

```

```

BASE COUNT 145 a 171 c 184 g 176 t
ORIGIN
alignment_scores:
Quality: 1062.00 Length: 225
Ratio: 4.872 Gaps: 3
Percent Similarity: 96.889 Percent Identity: 95.111

```

```

alignment_block:
US-09-776-865-2 x BE867611 ..
Align seg 1/1 to: BE867611 from: 1 to: 676

```

```

10 ArgProValGlnProAla..ArgProGlyGlyPheGlyLeuSerGlyArg 25
|||||
5 CGACCCACGGCGTCCGGCGCTCGCGGGGCTTGGGCTTGGGCGG 54
26 ArgSerLeuLeuGlyGlnValAlaSerThrProAlaHisValGlyValMe 42
|||||
55 CCGTCCCTCTCTCGCAGAGTGGCGAGTACACCTGCTCAGTACGATGAT 104
42 tArgSerProValAlaArgAspLeuAlaArgAsnAspGlyGlnLeuSerThr 59
105 GAGGTCTCCGGTTGAGACCTGGCCGGGAGACGATGGCAGAGACAGC 154
59 spArgThrProLeuLeuProGlyAlaProArgAlaGlyAlaAlaProVal 75
|||||
155 ACCGACGCGCTCTTCTACGGGCGGCCCGCCGGAAGCGCGCTCCAGT 204
76 CysCysSerAlaArgTyrAsnLeuAlaIleLeuAlaPhePheGlyPhe 92
|||||
205 TGTGCTCTGCTCGTTACACTAGCAATTTGGCCCTTTTGGTTCTT 254
92 eIleValTyrAlaLeuArgValAsnLeuSerValAlaLeuValAspMetV 109
|||||
255 CATGTGTATGCAATTCAGTGTGAATCTGAGTGTGCTTACGTGATATG 304
109 aAlaSerAsnThrThrLeuGlnAspAsnArgThrSerTyrAlaCysPro 125
305 TAGATTCAATACAACTTACAGATATATACACTTCCAAAGCATGTCCA 354
126 GluHisSerAlaProIleLeuValHisHisAsnGlnThrGlyLysTyr 142
|||||
355 GAGCATTCGCTCCCATTAAGATTTCATCATATCAAAACGGGTAGAAAGTA 404
142 rGlnThrPaspAlaGlnThrGlnGlyTyrIleLeuGlySerPhePhe.Tyr 158
405 CCAATGGATGCAGAACTCAAGATGGATTCCTGCTTCTTTTGGTAT 454
159 GlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerLysIle.. 174
|||||
455 GCGTACATCATCACACAGATTCCTGAGAGATATGTGGCACAACAAATTCG 504
175 GlyGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaValLeuThr 191
|||||
505 GGGGGGAAATGCTGCTGAGATTTGGGATCCTTGGCATCTGTCCTCAC 554
191 rLeuPheThrProIleAlaAlaAspLeuGlyValGlyProLeuIleVal 208
555 CCGTGTCTCCTCCATGCTGCGAGATTTAGAGATTGGACCATCTATTGTAC 604
208 euArgAlaLeuGlnGlyLeuGlyGlnGlyValThrPheProAlaMetHis 224
|||||
605 TCAGAGCAGCTAGAAAGAGACTAGAGAGGGGTGTACATTTCCAGCATGATT 654
225 AlaMetTyrSerSerTyrPala 231
655 GCCATGTGCTCTCTTGGGCT 675
seq_name: gb_est2:BI860521
seq_documentation_block: 736 bp mRNA linear EST 10-OCT-2001
LOCUS BI860521 603386787F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5395772 5',
DEFINITION mRNA sequence.
ACCESSION BI860521
VERSION BI860521.1 GI:16001268
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 736)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

```

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM12008 row: m column: 21
High quality sequence stop: 713.
Location/Qualifiers
1. 736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image:5395772"
/clone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 216 a 142 c 147 g 231 t
ORIGIN

alignment_scores:
Quality: 1024.50 Length: 224
Ratio: 5.022 Gaps: 2
Percent Similarity: 91.071 Percent Identity: 88.839

alignment_block:
US-09-776-865-2 x B1860521 ..

Align seg 1/1 to: B1860521 from: 1 to: 736

```

120 ThrSerLysAlaCysProGluHisSerAlaProIleLysValHisHis 136
1 ACTGCCAGGCGTCCAGACATTCGCTCCCTAAAGATTCATCTATA 50
136 ngInThGlyLysLysTyrGlnTrrPaspAlaGluThrGlnGlyTrrIle 153
51 TCAACGCGTAAGAAATACCAATGGATGCAGAAATCAAGAGATGATTC 100
153 euGlySerPhePheTyrGlyTrrIleLthrGlnIleProGlyTyr 169
101 TCGGTCCTTTTATGCTACATCATCACAGATTCCTGGAGATAT 150
170 ValAlaSerLysIleGlyLysMetLeuLeuGlyPheGlyIleLeuG 186
151 GTTCCAGCAAAATAGGGGGAAATCTCTAGGATTTGGGATCCTTGG 200
186 yThAlaValLeuThrLeuPheThrProIleAlaAlaAspLeuGlyVal 203
201 CACGTGCTCTCACCTGTTCACTCCATTCGTCAGATTAGAGATTG 250
203 LyrProLeuIleValLeuArgAlaLeuGlnGlyLeuGlyGlnGlyVal 219
251 GACCACCATTTGACTCAGACACTAGACAGACTAGAGAGGGGTGACA 300
220 PheProIaMetHisAlaMetTrrPseSerTrrPalaProProLeuG 236
301 TTTCACGCCATGCATGCGATGCTCTTCTTGGGCTCCCTCTTGAAG 350
236 gSerLysLeuLeuSerLysSerTrrAlaGlyAlaGlnLeuGlyThrVal 253
351 AAGCAAACTTCTAGCATTCATATGCAGAGACACAGCTTGGGACAGTAA 400
253 leSerLeuProLeuSerGlyIleLeuGlyTrrTyrMetLeuTrrPTh 269
401 TTTCTCTTCCCTTCTGGAATATTTGCTACTATATGAATGGAATTA 450

```

```

270 ValPheTrrPhePheGlyTrrIleGlyIlePheTrrPheLeuLeuTrrP 286
451 GTCTTCTACTT.TTTGCTACTATGGAATATTTGGTTCTTTCTGTGGAT 499
286 eTrrPLeuValSerAspThrProGlnLysHisLysArgIleSerHisTyr 303
500 CTGGTTAGTTAGTGACACACCACAAACACAGAGATTTCCCATATG 549
303 LuLysGlyTrrIleLeuSerSerLeuArgGlnGlnLeuSerSerGlnLys 319
550 AAAAGGATTCATTTTCATTCATTAAGAAATCAG..... 584
320 SerValProTrrPValProIleLeuLysSerLeuProLeuTrrPalaI 336
585 .....GTATGCACTTGG.. 596
336 lValAlaHisPheSerTrrAsn 343
597 .....ACACATTTTAAATATAT 614
seq_name: gb_est2:BF676817
seq_documentation_block:
LOCUS BF676817 626 bp mRNA linear EST 21-DEC-2000
DEFINITION 602084380F1 NIH_MGC_83 Homo sapiens CDNA clone IMAGE:4248811 5',
mRNA sequence.
ACCESSION BF676817
VERSION BF676817.1 GI:11950712
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 626)
NIH-MGC http://mhc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHCMI067 row: o column: 20
High quality sequence stop: 598.
Location/Qualifiers
1. 626
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image:4248811"
/clone_lib="NIH_MGC_83"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggccgctcgagc); Site_2: SfiI (ggccattatgac
); 5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGCGC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGCGCGAGCATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT 157 a 126 c 123 g 220 t
ORIGIN

```

alignment_scores:
Quality: 991.00 Length: 189
Ratio: 5.243 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-776-865-2 x BF676817 ..

Align seg 1/1 to: BF676817 from: 1 to: 626

```

348 ThrLeuLeuThrLeuLeuProThrTyMetLysGluIleLeuArgPheas 364
|||||
2 ACTTATTGACCTATTGCTGCTACTATTATGAAGAGATCCTAGGTTCAA 51
|||||
364 nValIngluAsnGlyPheLeuSerSerLeuProThrLeuGlySerTrpL 381
|||||
52 TGTTCAGAGAGATGGGTTTTTATCTTCATGCTGCTATTAGGCTCTGGT 101
|||||
381 euCyMetIleLeuSerGlyGlnAlaAlaAspAsnLeuArgAlaLysTrp 397
|||||
102 TATGATATGATCTGTCTGCTCAAGCTGCTGACAAATTTAAGGCAAAATGG 151
|||||
398 AsnPheSerThrLeuCySerValArgArgIlePheSerLeuIleGlyMetI 414
|||||
152 AATTTTCAACTTATGTTGTTGCGAGATTTTACCTTATAGAGATGAT 201
|||||
414 eGlyProAlaValPheLeuValAlaAlaGlyPheIleGlyCysAspTyrS 431
|||||
202 TGGACCTCAGATATCTGTGTAGCTGTGCTTCATGTGCTGTGATATAT 251
|||||
431 erLeuAlaValAlaPheLeuThrIleSerThrThrLeuGlyGlyPheCys 447
|||||
252 CTTTGCCGCTGCTTCTCACTATATATCAACAACACTGGAGGCTTTTGGC 301
|||||
448 SerSerGlyPheSerIleAsnHisLeuAspIleAlaProSerTyrAlaG 464
|||||
302 TCTTCTGATTTAGCATCAACATCTGATATGCTCTGCTGATGCGG 351
|||||
464 YIleLeuLeuGlyIleThrAsnThrPheAlaThrIleProGlyMetValG 481
|||||
352 TATCTCTCTGGGCATCAACAATATACATTTGCCATATTCAGAAATGGTTG 401
|||||
481 YProValIleAlaIleAsnSerLeuThrProAspAsnThrValGlyGlyL 497
|||||
402 GGCCCGCATTTGCTAAAGTCTGACCCCTGATTAACACATGTGGAGAAATGG 451
|||||
498 GlnThrValPheTyrIleAlaAlaIleAsnValPheGlyAlaIlePhe 514
|||||
452 CAACCGGTCTCTATATGCTGCTCTATTAATGTTTTTGGCCATTTT 501
|||||
514 ePheThrLeuPheAlaLysGlyGluValGlnAsnTrpAlaLeuAsnAspH 531
|||||
502 CTTTACACTATTCGCCCAAGGTGAGTACAAAACTGGGCTCTCATATGATC 551
|||||
531 IShISGLYHISARGHIS 536
|||||
552 ACCATGACACGACGAC 568

```

seq_name: gb_est2:BI697765

seq_documentation_block:

LOCUS BI697765 711 bp mRNA linear EST 18-SEP-2001

DEFINITION 603346838F1 NCI_CGAP_Mam2 Mus musculus cDNA IMAGE:5374407 5', mRNA sequence.

ACCESSION BI697765

VERSION BI697765

KEYWORDS BI697765.1 GI:15660394

SOURCE EST.

ORGANISM house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 711)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

plate: L1AM1953 row: c column: 16

High quality sequence stop: 707.

FEATURES

source

Location/Qualifiers

1..711

/organism="Mus musculus"

/strain="FVB/N-3"

/db_xref="taxon:10090"

/clone="IMAGE:5374407"

/clone_lib="NCI_CGAP_Mam2"

/tissue_type="tumor, biopsy sample"

/dev_stage="5 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1; Salt: Site:2; Notti; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

BASE COUNT

147 a 191 c 199 g 174 t

alignment_scores:

Quality: 991.00 Length: 236

Ratio: 4.546 Gaps: 2

Percent Similarity: 92.373 Percent Identity: 83.051

alignment_block:

US-09-776-865-2 x BI697765 ..

Align seg 1/1 to: BI697765 from: 1 to: 711

```

33 AlaserThrProAlaHis.Val.....GlyValMetArgSerProValA 47
|||||
6 GCACGACGCGCGCACATCTCTCCAGCGGTGTGAGCCCTGCTTC 55
|||||
47 TAspLeuAlaArgAsnAspGlyGluGluSerThrAspArgThrProLeu 63
|||||
56 GGGGTCCGCGCGGAAACAGCATGAGGAGACTCGAGACAGACCCGCTC 105
|||||
64 LeuProGlyAlaProArgAlaGluAlaAlaProValCysSerSerAlaAr 80
|||||
106 CTGCCGCGCGCGCGGACAGCCGAGCGGCTCCAGTGTGCTCTGCTGCTG 155
|||||
80 GTTAsnLeuAlaIleLeuAlaPhePheGlyPhePheIleValTyrAla 97
|||||
156 GTACACTTACGATTTTGGCGTCTGTGCTCTGCTCTCTATGCT 205
|||||
97 euArgValAsnLeuSerValAlaLeuValAspMetValAspSerAsnThr 113
|||||
206 TACGGGTAACTGATGCTTCTGCTTACGTAGGACATGATGATTTCAATCA 255
|||||
114 ThrLeuGluAspAsnArgThrSerLysAlaCysProGluHisSerAlaPr 130
|||||
256 ACTCTGACTGATTAATAGAACGCTTAAGAGAGTGTCCGGAACATTCGCC 305
|||||
130 oILeLysValHisHisAsnGlnThrGlyLysTyrGlnTrpAspAlaG 147
|||||
306 CATAAAGTTCACCAATCACACAGGTAAAGATACAGATGGAGTGCAG 355
|||||
147 IuThrGlnGlyTrpIleLeuGlySerPhePheTyrGlyTyrIleIleThr 163
|||||
356 AAACCTCAAGGTGATCTGCGCTC.TTTTTCACGCTACATGCTCAC 404
|||||
164 GlnIleProGlyGlyTyrValAlaSerLysIleGlyLysMetLeuLe 180
|||||
405 CAGATTCGCCGCTGGGTACATGTCAGCAGGCTCGAGGAGGAAGCTGCTG 454

```


VERSION BF124137.1 GI:10963177
 EST.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 904)
 AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabds-remail.nih.gov
 Tissue Procurement: Lotmar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LHAM9283 row: k column: 09
 High quality sequence stop: 680.

FEATURES
 source location/Qualifiers
 1..904
 /organism="Mus musculus"
 /strain="C57/B6"
 /db_xref="taxon:10090"
 /clone="IMAGE:4024064"
 /clone_lib="NCL-GAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lotmar Hennighausen/Robin Humphreys, NIH"

BASE COUNT 193 a 214 c 276 g 221 t
 ORIGIN

alignment_scores:
 Quality: 909.00 Length: 233
 Ratio: 4.170 Gaps: 0
 Percent Similarity: 93.562 Percent Identity: 84.549

alignment_block:
 us-09-776-865-2 x BF124137 ..

Align seg 1/1 to: BF124137 from: 1 to: 904

40 GYValMetArgSerProValArgAspLeuAlaArgAsnAspGlyGluG1 56
 |||||
 24 GCGGTATGAGGCCCTCGTCCGGGTCGGCGGGAACGACGATGAGCA 73
 |||||
 56 userThrasparGthrProLeuLeuProGlyAlaProArgAlaGluAlaA 73
 |||||
 74 GAGCTGGAGACAGACCCGCTCTCGCGCGCGGACGACGACGACGCG 123
 |||||
 73 laProValCysCysSerAlaArgTyrAsnLeuAlaIleLeuAlaPhePhe 89
 |||||
 124 CTCGAGTGTGCTGCTCTCGGTACCACTTACGATTTGGCGTTCTGT 172
 |||||
 90 GlyPhePheIleValTyrAlaLeuArgValAsnLeuSerValAlaLeuVa 106
 |||||
 173 GGTTCCTTCGTTCTATGCTTACGGGTGAACCTGAGTGTGGCTTAGT 221
 |||||
 106 lAspMetValaspSerAsnThrLeuGluAspAsnArgThrSerLysA 123
 |||||
 222 GGACATGTAGATTCAATACACTGACGTGATATAGAACGCTTAGAG 271
 |||||
 123 lacysProGluHisSerAlaProIleLysValHisHisGlnThrGly 139
 |||||
 272 AGTGTGCGGAACATTTCTGCCCCCATTAAGTTACACATACACAGGT 321
 |||||

140 LysLysTyrGlnTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPh 156
 |||||
 322 AAAAGTACAGATGGATGACAGAAACTCAAGGTCGATTCCTGCTCTT 371
 |||||
 156 ePheTyrGlyTyrIleIlePheGlnIleProGlyGlyTyrValAlaSerL 173
 |||||
 372 TTTTACCGGTACATCGTCACCCAGATTCCTCGGTGGTACATTTGCCACA 421
 |||||
 173 ySileGlyGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaVal 189
 |||||
 422 GGGTGGAGGGAAGCTGCTGGGCTGGGCTGATCTTAGCACTTCCTC 471
 |||||
 190 LeuThrLeuPheThrProIleAlaAlaAspLeuGlyValGlyProLeuI 206
 |||||
 472 TTCACCCGTTCACACCGCTGGCCGACACTTAGCGGTGACGCTCTGT 521
 |||||
 206 eValLeuArgAlaLeuGluGlyLeuGlyGlyValThrPheProAlaM 223
 |||||
 522 TGTGCTTAGACGCTGGAAGACTGGAGAGGTGTACGTTTCCACGCTA 571
 |||||
 572 TGACAGC.ATGTGTCTTCCTGGGCTCCCTCGGAAGAAGCAAGCTT 620
 |||||
 223 eThiSalMetTrpSerTrpAlaProProLeuGluArgSerLysLeu 239
 |||||
 240 LeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrValIleSerLeuP 256
 |||||
 621 CTTCACATTTCTCTATGCGGAGACACAGTT.GGACACAGATCTCATTCC 669
 |||||
 256 oLeuSerGlyIleIleCysTyrTyrMetAsnTrpThrTyrValPheTyr 272
 |||||
 670 TCTTTC.GGAATTAATGTACTATATGAATGACCTTACCTCTCTCAT 717
 |||||

seq.name: gb_hlc:BC023331

seq_documentation_block:
 LOCUS BC023331 1531 bp mRNA linear HTC 05-FEB-2002
 DEFINITION Mus musculus, clone IMAGE:450038, mRNA.
 ACCESSION BC023331
 VERSION BC023331.1 GI:18497348
 KEYWORDS HTC.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 1531)
 AUTHORS Strausberg, R.
 DIRECT SUBMISSION
 SUBMITTED (05-FEB-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 CONTACT: MGC help desk
 Email: cgabds-remail.nih.gov
 Tissue Procurement: The Cepko Laboratory
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: angbcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

REMARK
 COMMENT
 JOURNAL

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 29 Row: f Column: 13
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis

ACCESSION BG541099
 VERSION BG541099.1 GI:13533332
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 893)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-femail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LCM1519 row: a column: 17
 High quality sequence stop: 549.
 Location/Qualifiers
 1..893
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4694680"
 /clone_lib="NIH-MGC-77"
 /lab_host="DH10B (rT1 phage-resistant)"
 /note="Organ: Lung; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggcgctcgcc); Site_2: SfiI (ggcattatggc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
 5'-ATCTGAGGCGGCGGCCGACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.9
 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH-MGC Library."
 BASE COUNT 236 a 169 c 214 g 274 t
 ORIGIN
 alignment_scores:
 Quality: 900.00 Length: 325
 Ratio: 4.072 Gaps: 9
 Percent Similarity: 68.000 Percent Identity: 63.077
 alignment_block:
 US-09-776-865-2 x BG541099 ..
 Align seg 1/1 to: BG541099 from: 1 to: 893
 203 GlyProLeuIleValIleuArgAlaIleuGluGlyGluGlyValTh 219
 1 GGCACCACTGTTAGCATTCATATGAGAGACAGAGGCTGTAC 50
 219 rPheProAlaMetHisAlaMetTrpSerSerTrpAlaProProLeuGlnA 236
 51 ATTTCACACCATGCATGCATGTGTCTTCTTGCGCTCCCTTTGAAA 100
 236 rGSerLysLeuLeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrVal 252
 101 GAACCAAACTCTTAGCATTTTCATATGAGAGACAGAGCTGGACAGTA 150
 253 IleSerLeuProLeuSerGlyIleIleCysTyrTrpMetAsnTrpThrTy 269
 151 ATTTCCTCTCTCTTCTGGAATATTGCTACTATATGAAATTGCACTTA 200
 269 rValPheTyrPhePheGlyThrIleGlyIlePheTrpPheLeuLeuTrp 285
 201 TGTCTTACTT.TTGTGACTATTGGAATATTGTTCTTTG...TGG 246

286 IleTrpLeuValSerAspThrProGlnLysHisLysArgIleSerHisTy 302
 247 ATCGGTAGTATTAGACACACACCAAAACACAGAAATTCACATA 296
 302 rGlnLysGluTyrIleLeuSerSerLeuArgAsnGlnLeuSerSerGlnL 319
 297 TGAAGAAGCAATACATTTCTTCATCATTAAGAAATCAG..... 333
 319 ySerValProTrpValProIleLeuLysSerLeuProLeuTrpAlaIle 335
 333 333
 336 ValValAlaHisPheSerTyrAsnTrpThrPheTyrThrLeuThrIle 352
 333 333
 352 uLeuProThrTyrMetLysGluIleLeuArgPheAsnValGlnGlnAsnG 369
 334AATG 337
 369 LypheLeuSerSerLeuProTyrLeuGlySerTrpLeuCysMetIleLeu 385
 338 GGTTTTATCTTCATTCCTTATTTAGCCTTGTGTTATGATATCATCTG 387
 386 SerGlyGlnAlaAlaAspAsnLeuArgAlaLysTrpAsnPheSerThrL 402
 388 TCTGTCCAACTCTGTCACATTTAAGGCAAAATGCAATTTTCAACT 437
 402 eucGValArgArgIlePheSerLeuIleGlyMetIleGlyProAlaVal 418
 438 TATGTGTCCGAGAAATTTAGCCTTATAGCAATGATTTGACCTGCAGTA 487
 419 PheLeuValAlaAlaGlyPheIleGlyCysAspTyrSerLeuAlaVal 435
 488 TTCTGTGAGCTGCTGGCTTCATTTGGTGATATTTCTTTGGCCGTCT 537
 435 aPheLeuThrIleSerThrLeuGlyGlyPheCysSerSerGlyPheS 452
 538 TTCTAATATATCAACAACACT...GGAGGCTTTGGTCTCTTCGATTTA 584
 452 erIleAsnHisLeuAspIleAlaProSerTyrAlaGlyIleLeuLeuG 468
 585 GATCCACCATTTGGGATATG..... 605
 468 yIleThrAsnThrPheAlaThrIleProGlyMetValGlyProValIleA 485
 606GTCCTCCACACCTTTGGGAAATGCGAA 636
 485 IalysSerLeuThrProAspAsnThrValGlyLysTrpGlnThrValPhe 501
 637 CCGGCTCTCTT.....ATGGTGGGGGAAATTAAGTCT..... 668
 502 TyrIleAlaAlaAlaIleAsnValPheGlyAlaIlePheThrLeuPh 518
 669GGTGGCCACTTTCATTCACATTT 691
 518 eAlaLysGlyLysValGln 524
 692 TGGCAAAAGTGAGGCTCCAA 710
 seq_name: gb_est1:BB613552
 seq_documentation_block:
 LOCUS BB613552 682 bp mRNA linear EST 26-OCT-2001
 DEFINITION BB613552 RIKEN full-length enriched, 10 day neonate skin Mus
 ACCESSION BB613552
 VERSION BB613552.1 GI:16454120
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

| | | | | | |
|------------|--|---|---------------------------------|----------------------|-------------------------------------|
| LOCUS | BB537525 | 645 bp | mRNA | linear | EST 26-Oct-2001 |
| DEFINITION | BB537525 | RIKEN full-length cDNA, 0 day neonate eyeball | Mus musculus | CDNA clone EJ3006F23 | 3' similar to AJ387747 Homo sapiens |
| ACCESSION | BB537525 | BB537525 | mRNA for stalin, mRNA sequence. | | |
| VERSION | BB537525.2 | GI:16446752 | | | |
| KEYWORDS | EST. | | | | |
| SOURCE | house mouse. | | | | |
| ORGANISM | Mus musculus. | | | | |
| REFERENCE | Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 645) | | | | |
| AUTHORS | Arakawa,T., Carolin,I., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Ozakaki,S., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasakii,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y. | | | | |
| TITLE | RIKEN Mouse ESTs (Arakawa,T., et al. 2001) | | | | |
| JOURNAL | Unpublished (2001) | | | | |
| COMMENT | On Jul 31, 2000 this sequence version replaced gi:9593025. | | | | |

seq_name: gb_est1:AL597124

seq_documentation_block:

LOCUS AL597124 499 bp mRNA linear EST 14-AUG-2001

DEFINITION DKFZp313H1911.t1 313 (synonym: hlcc2) Homo sapiens cDNA clone

ACCESSION AL597124

VERSION AL597124.1 GI:15154820

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 499)

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE Koehler, K., Beyer, A., Mewes, W., Weill, B. and Wiemann, S.

JOURNAL EST (Koehler, K., Beyer, A., Mewes, W., Weill, B. and Wiemann, S.)

COMMENT Unpublished (1999)

CONTACT: Koehler K

MIPS

Am Klopferstr. 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by BMFZ (Biomedical Research Center at the Charité, Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.

No sl sequence available.

This clone (DKFZp313H1911) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY. Email: clone@rzpd.de.

Location/Qualifiers

1. 499

source

1. /organism="Homo sapiens"

2. /db_xref="taxon:9606"

3. /clone="DKFZp313H1911"

4. /clone_1ib="313 (synonym: hlcc2)"

5. /dev_stage="adult"

6. /lab_host="DH10B"

7. /note="Vector: pTRIPlex2; Site_1: SfiIA; Site_2: SfiIB; cDNA-collection"

BASE COUNT 110 a 129 c 133 g 126 t 1 others

ORIGIN

alignment_scores:

Quality: 862.00 Length: 165

Ratio: 5.256 Gaps: 0

Percent Similarity: 99.394 Percent Identity: 98.182

Alignment_block:

US-09-776-865-2 x AL597124 ..

Align seg 1/1 to: AL597124 from: 1 to: 499

9 ProArgProValGlnProAlaArgProGlyGlyPheGlyLeuSerGlyArg 25

3 CCCCAGCCCGGTCAGACAGCTGCGGCGGCTTGGCGCTGTCGGGCGG 52

25 GAGSerLeuLeucysGlnValAlaSerThrProAlaHisValGlyValM 42

53 GCGCTCCCTCTCTGCGAGGTGGAGTACACCTGCTCAGTAGCGCTCA 102

42 eArgSerProValArgAspLeuAlaArgAsnAspGlyGluGluSerThr 58

103 TGAAGTCTCGGTTCCGAGACCTGCGCCGGAACGATGGCGAGAGACACG 152

59 AspArgThrProLeuLeuProGlyAlaProArgAlaGluAlaAlaProVa 75

153 GACCGCAGCGCTTCTTACCGGGCGCCCAAGGCGGCGCTCCAGT 202

75 lGycGysSerAlaArgTyrAsnLeuAlaIleLeuAlaPhePheGlyPhe 92

203 GTGCTGCTGCTGCTGCTTACACCTTAGAATTTTGCCCTTTTGGTTCCT 252

92 heIleValTYrAlaLeuArgValAsnLeuSerValAlaLeuValAspMet 108

253 TCATTTGTTGATGCTATTGAGTCTTATCTGAGCTGTGGCTTAGTGATATG 302

109 ValAspSerAsnThrThrLeuGluAspAsnArgThrSerIlysalAcysPr 125

303 GTAAATTCAAATACACCTTGAAGATTAATGAACTTCCAAAGCGGTGCC 352

125 oGluHisSerAlaProIleLysValHisHisArgGlnThrGlyLysIst 142

353 AGACCAATTCATCTCCATTAAGATTCAATCAATCAACAGGGTAAAGAT 402

142 YrGlnTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTyr 158

403 ACCAATGGAGTGCAGAACTCAAGATGATTCGGTTCCTTTTAT 452

159 GLYTYrIleIleThrGlnIleProGlyGlyTYrValAlaSerLys 173

453 GCGTACATMTATCACAGATTCCTGAGAGATGCTTGCACAGCAA 497

seq_name: gb_est1:AA833297

seq_documentation_block:

LOCUS AA833297 518 bp mRNA linear EST 23-FEB-1998

DEFINITION ud05d09.r1 Soares, NMPu Mus musculus cDNA clone IMAGE:1434257 5' similar to WP:C36C10.2 CE00105 SODIUM/PHOSPHATE TRANSPORTER ;, mRNA sequence.

ACCESSION AA833297

VERSION AA833297.1 GI:2907025

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 518)

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LINTL; contact the IMAGE Consortium (info@image.lnlnl.gov) for further information.

Seq primer: -28m13 rev2 ET from Amerisham

MGI:918325

High quality sequence stop: 444.

Location/Qualifiers

1. 518

source

1. /organism="Mus musculus"

2. /db_xref="taxon:10090"

3. /clone="IMAGE:1434257"

4. /clone_1ib="Soares_NMPu"

5. /sex="female"

6. /dev_stage="adult"

7. /lab_host="DH10B"

8. /note="Organ: uterus; Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pregnant mouse uterus, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTR73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 123 a 105 c 121 g 169 t
ORIGIN

alignment_scores:
Quality: 802.00 Length: 172
Ratio: 4.981 Gaps: 0
Percent Similarity: 93.605 Percent Identity: 85.465

alignment_block:
US-09-776-865-2 x AA833297 ..

Align seg 1/1 to: AA833297 from: 1 to: 518

```
275 GlyThrIleGlyIlePheTrpPheLeuLeuTrpIleTrpLeuValSerAs 291
||| ::::||| ||||| ::::||| ::::||| ::::|||
1 GGTATAGCTGGAAATGTCTGCTTTATTTATGATGTGATAGTCAGTCA 50
291 pThrProGlnLysHisLysArgIleSerHisTrpGlnLysGlnTrpIleL 308
||||| :::: ||||| ||||| ||||| ||||| |||||
51 TACACCAGAACTCACAGACAACTCCCATATGAAAAAGAAATACATTG 100
308 euSerSerLeuArgAsnGlnLeuSerSerGlnLysSerValProTrpVal 324
:::||||| ::::||| ||||| ||||| ||||| |||||
101 TTTCAATCAATTAATAATCAGCTTCTTCGACAGAGGTGGCTGGGGG 150
325 ProIleLeuLysSerLeuProLeuTrpAlaIleValAlaHisPheSe 341
||||| ||||| ||||| ||||| ||||| ||||| |||||
151 TCCATCTTGAACTCAGCTGCACCTTGGGCAATTGGTAGCACATTCTC 200
341 rTyrAsnTrpIlePheTrpThrLeuLeuThrLeuLeuProThrTyrMetL 358
||||| ::::||| ||||| ||||| ||||| ||||| |||||
201 CTACAACTGGTCTTTTACACCTTATGTAGCGCTACTGCCAACTATATGA 250
358 yGlnIleLeuArgPheAsnValGlnGluAsnGlyPheLeuSerSerLeu 374
||||| ||||| ||||| ||||| ||||| ||||| |||||
251 AGGAAATCCTAAGGTTCATGTCAAGAGAAAGGGGTTTATCTGCATTG 300
375 ProTyrLeuGlySerTrpLeuLysMetIleLeuSerGlyGlnAlaAlaAs 391
||||| ::::||| ||||| ||||| ||||| ||||| |||||
301 CCTTATTTGGCTGTGGTATGATGATCCTCTGTGGTCAAGCCGCTGA 350
391 pAsnLeuArgAlaLysTrpAsnPheSerThrLeuLysValArgArgIleP 408
| ||||| ::::||| ||||| ||||| ||||| ||||| |||||
351 CTATTTAAGGTCAGTGAAGTCAACTTTCAACTATAGCGTTGCGAGCATTT 400
408 heSerLeuIleGlyMetIleGlyProAlaValPheLeuValAlaIleGly 424
||||| ::::||| ||||| ||||| ||||| ||||| |||||
401 TTAGCCTCGTAGGAATGTTGGCCTGGGTTTCTTAGTCGGGCTGGA 450
425 PheIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerTh 441
||||| ||||| ||||| ||||| ||||| ||||| |||||
451 TTTATAGGCTGTGACTATTCCTTGSCCGTTGCGTTCTTACCATATCCAC 500
441 rThrLeuGlyGlyPhe 446
||||| ||||| ||||| ||||| ||||| ||||| |||||
501 GACGCTGGAGGCTTC 516
```


Tue Jul 16 08:11:13 2002

OM of: US-09-776-865-2 to: GenEmbl:* out-format : pfs
Date: Jul 15, 2002 3:35 PM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=frame+ p2n.model -DEV=x1h
-O=/cgn2.1/USPRO.spool/US09776865/runat_15072002.082208.27415/app_query.fasta_1.1159
-DB=GenEmbl -OPMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
-GAPX4=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=Dlosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTTRF=pfs
-NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-DEV=US09776865_@CGN1_1.5372 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-776-865-2

Query length: 536

Database: GenEmbl:*

Database sequences: 1797656

Database length: 187333701

Search time (sec): 3807.270000

score_list:

| Sequence | Strid | Orig | ZScore | Escore | Len | Documentation |
|-----------------|-------|---------|---------|----------|--------|---------------------------------|
| gb_pat:AX207624 | + | 2836.00 | 4141.55 | 2.5e-222 | 2930 | AX207624 Sequence 1 from Patent |
| gb_pr:AF244577 | + | 2836.00 | 4141.55 | 2.5e-222 | 2930 | AF244577 Homo sapiens membrane |
| gb_pat:AX138494 | + | 2804.00 | 4096.26 | 8.4e-220 | 2512 | AX138494 Sequence 2 from Patent |
| gb_pr:HA387747 | + | 2804.00 | 4096.26 | 8.4e-220 | 2512 | AJ387747 Homo sapiens mRNA for |
| gb_pr:BC020961 | + | 2770.50 | 4043.92 | 6.9e-217 | 3329 | BC020961 Homo sapiens, soluble |
| gb_com:AF244578 | + | 2338.00 | 3410.44 | 1.3e-181 | 2844 | AF244578 Ovis aries membrane g |
| gb_pat:AX207626 | + | 2338.00 | 3410.44 | 1.3e-181 | 2844 | AX207626 Sequence 3 from Patent |
| gb_in:AY060776 | + | 992.00 | 1438.35 | 9.4e-72 | 1841 | AY060776 Drosophila melanogast |
| gb_hlg:AC014246 | + | 984.50 | 1395.08 | 2.4e-69 | 3344 | AC014246 Drosophila melanogast |
| gb_in:AC023665 | + | 984.50 | 1377.13 | 2.4e-68 | 167928 | AC023665 Drosophila melanog |
| gb_in:AC023711 | + | 984.50 | 1369.84 | 2.5e-68 | 174163 | AC023711 Drosophila melanog |
| gb_in:AE003491 | + | 984.50 | 1369.84 | 6.1e-68 | 323461 | AE003491 Drosophila melanog |
| gb_ro:AF271235 | + | 973.00 | 1401.86 | 1.0e-69 | 3982 | AF271235 Rattus norvegicus dif |
| gb_ro:AF24864 | + | 971.00 | 1403.98 | 7.7e-70 | 2528 | AF24864 Mus musculus vesicula |
| gb_pr:AB032435 | + | 968.00 | 1394.62 | 2.6e-69 | 3346 | AB032435 Homo sapiens DNPI mR |
| gb_pat:IA0028 | + | 942.00 | 1362.86 | 1.5e-67 | 2716 | IA0028 Sequence 1 from Patent |
| gb_pat:IA0029 | + | 942.00 | 1360.59 | 2.0e-67 | 2716 | IA0029 Sequence 3 from Patent |
| gb_pat:IA0220 | + | 942.00 | 1360.59 | 2.0e-67 | 2716 | IA0220 Sequence 1 from Patent |
| gb_pat:IA0221 | + | 942.00 | 1360.59 | 2.0e-67 | 2716 | IA0221 Sequence 3 from Patent |
| gb_pat:IA0222 | + | 942.00 | 1360.59 | 2.0e-67 | 2716 | IA0222 Sequence 1 from Patent |
| gb_pat:IA0223 | + | 942.00 | 1360.59 | 2.0e-67 | 2716 | IA0223 Sequence 3 from Patent |
| gb_ro:RND07609 | + | 939.00 | 1359.45 | 2.3e-67 | 2024 | U07609 Rattus norvegicus brain |
| gb_in:AY069501 | + | 928.50 | 1340.63 | 2.6e-66 | 2748 | AY069501 Drosophila melanogast |
| gb_in:CE038C10 | + | 906.00 | 1279.54 | 6.6e-63 | 34193 | Z19153 Caenorhabditis elegans |
| gb_pr:AK025880 | + | 894.00 | 1296.06 | 7.9e-64 | 1587 | AK025880 Homo sapiens cDNA: FL |
| gb_ro:BC018306 | + | 893.50 | 1293.88 | 1.1e-63 | 1874 | BC018306 Mus musculus, similar |
| gb_hlg:AC020077 | + | 891.00 | 1253.40 | 1.8e-61 | 47723 | AC020077 Drosophila melanog |
| gb_in:AC091501 | + | 891.00 | 1237.95 | 1.4e-60 | 198390 | AC091501 Drosophila melanog |
| gb_in:AE003806 | + | 891.00 | 1234.99 | 2.1e-60 | 268219 | AE003806 Drosophila melanog |
| gb_in:AF024691 | + | 886.50 | 1264.91 | 4.3e-62 | 9703 | AF024691 Drosophila ananassae |
| gb_pr:AB020527 | + | 876.00 | 1264.03 | 4.1e-62 | 2626 | AB020527 Homo sapiens mRNA for |
| gb_pr:AK024903 | + | 867.00 | 1247.25 | 4.1e-61 | 3616 | AK024903 Homo sapiens cDNA: FL |
| gb_in:AC023715 | + | 856.00 | 1188.89 | 7.4e-58 | 160622 | AC023715 Drosophila melanog |
| gb_in:AE003432 | + | 856.00 | 1181.24 | 2.0e-57 | 319551 | AE003432 Drosophila melanog |
| gb_hlg:AC013869 | + | 840.00 | 1194.42 | 3.5e-58 | 11568 | AC013869 Drosophila melanog |
| gb_in:AY075277 | + | 839.50 | 1125.19 | 2.5e-59 | 1710 | AY075277 Drosophila melanog |
| gb_hlg:AC014600 | + | 839.50 | 1181.85 | 1.8e-57 | 34229 | AC014600 Drosophila melanog |
| gb_in:AE003640 | + | 839.50 | 1158.98 | 4.4e-56 | 267488 | AE003640 Drosophila melanog |
| gb_in:DROSADH01 | - | 839.50 | 1157.57 | 4.1e-56 | 303885 | AE003407 Drosophila melanog |

| | | | | | | |
|-----------------|---|--------|---------|---------|--------|-----------------------------|
| gb_hlg:AC020252 | + | 837.00 | 1177.36 | 3.2e-57 | 36842 | AC020252 Drosophila melano |
| gb_in:AC007756 | - | 837.00 | 1160.35 | 2.9e-56 | 170105 | AC007756 Drosophila melan |
| gb_in:AC007757 | - | 837.00 | 1160.29 | 2.9e-56 | 170939 | AC007757 Drosophila melan |
| gb_in:AE003730 | - | 837.00 | 1157.42 | 4.2e-56 | 221409 | AE003730 Drosophila melan |
| gb_com:RABRCCTX | + | 829.00 | 1198.86 | 2.1e-58 | 1855 | M76466 Rabbit renal cortica |

seq_name: gb_pat:AX207624

seq_documentation_block:

LOCUS AX207624 2930 bp DNA linear PAT 31-AUG-2001

DEFINITION Sequence 1 from Patent WO01565398.

ACCESSION AX207624

VERSION AX207624.1 GI:15422329

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

ORIGIN

BASE COUNT

alignment_scores:

Quality: 2836.00

Ratio: 5.291

Percent Similarity: 100.000

Percent Identity: 100.000

alignment_block:

US-09-776-865-2 x AX207624

Align seg 1/1 to: AX207624 from: 1 to: 2930

1 MetAlaAlaGlyAlaMetThrProProArgProValGlnProAlaArgTr 17

|||||

263 ATGGCGGGGGGGGCGAGGACCGCCCGCGGTCACGACGTCGCGC 312

|||||

17 OGtYGLYPheGlyLeuSerGlyArgSerLeuLeuGlyValAlaAla 34

|||||

313 CGGGGGCTTGGCGCTGCGGGCGCGCTCTCTCTGCGAGGTGGCA 362

|||||

34 eThrProAlaHisValGlyValMetArgSerProValArgAspLeuAla 50

|||||

363 GTCACCTGCTCGTACGTCATGAGTCGCGGTTCGAGACCTGGCC 412

|||||

51 ArgAsnAspGlyGluGluSerThrAspArgThrProLeuLeuProGlyAl 67

|||||

413 CGGAGCAGTGGCGAGGAGGACGACCGCGCTCTCTTACCGGCGCG 462

|||||

67 aProRgaLagLualAaLProValCysCysSerAlaArgTyraSLeu 84
 443 CCCAGGGGCGAAGCGCGTCCAGTGTGCTGCTGCTTACCACTTAG 512
 84 LaLeLeuAlaPhePheGlyPhePheLeValTyraLalaLeuArgValasn 100
 513 CAATTGTGGCCCTTTTGTGTTCTTCATTTGTATGATGATGATGATGAT 562
 101 LeuSerValAlaLeuValAspMetValAspSerAsnThrThrLeuGlnAs 117
 563 CTGAGTGTGGCTTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 612
 117 PasaRgThrSerLysAlaCysProGlnHisSerAlaProLeuLysValH 134
 613 TAAATAGACTTCCAAAGCGGTGTCAGAGCATCTGCTCCCATTAAGTTC 662
 134 LSHASnGlnThrGlyLysLysTyrglnThrPaspAlaGluThrGlnGly 150
 663 ATCAATATCAAAAGGGTAAGAGTACCAATGGATGACAAAGCTCAGAGA 712
 151 TrPileLeuGlySerPhePheTyrglyTyrlleIleThnGlnIleProgl 167
 713 TGGATTCTCGGTTCTTTTATATGCTACATCATCATCAGACAGATTCCTG 762
 167 yGlyTyraValaLaserLysIleGlyLysMetLeuLeuGlyPheGlyI 184
 763 AGGATATGTGTGCCAGCAAAATAGGGGGGAAAATGCTGCTAGATTTGGGA 812
 184 LeuGlyThrAlaValLeuThrLeuPheThrProIleAlaAlaLaspLeu 200
 813 TCTTGGGACGTGCTGTCTCACCTGTTCTACCTCCATTCGTGACATTTA 862
 201 GlyValGlyProLeuIleValLeuArgAlaLeuGlnGlyLeuGlyGluI 217
 863 GGAGTTGGACCACTCATTTGTACTGACAGCAGTACAGAGACTAGAGAGG 912
 217 yValThrPheProAlaMetHisAlaMetTrpSerSerTrpAlaProL 234
 913 TGTATCATTTTCAGCCATGATGATGATGATGATGATGATGATGATGAT 962
 234 euGlnArgSerLysLeuSerLysLeuSerLysLeuSerLysLeuSerLys 250
 963 TTGAAGAAGCAAACTCTTACATTTGATGATGATGATGATGATGATGAT 1012
 251 ThrValIleSerLeuProLeuSerGlyIleIleCysTyTyMetAsnTr 267
 1013 ACAGTAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1062
 267 pThrTyraValPheTyPhePheGlyThrIleGlyIlePheThrPheLeu 284
 1063 GACTTATGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1112
 284 euTrpIleTrpLeuValSerAspThrProGlnLysHisLysArgLleSer 300
 1113 TGTGATGTGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 1162
 301 HisTyrgLysGluTyrlleLeuSerSerLeuArgAsnGlnLeuSerSe 317
 1163 CATATATGAAGAAGATACATTTCTTTCATCATTAAGAAATCAGCTTTCTT 1212
 317 rGlnLysSerValProTyPValProIleLeuLysSerLeuProLeuTrpA 334
 1213 ACAGAAAGTCACTGCGGTGTACCATTTTAAATCCCTGACACTTTGGG 1262
 334 laIleValValAlaHisPheSerTyraAsnTrpThrPheTyThrLeuLeu 350
 1263 CTATCGTGTGGTGCACACTTTCTTACCACTGACACTTTTAACTTATTTG 1312
 351 ThrLeuLeuProThrTyMetLysGluIleLeuArgPheAsnValGlnG 367
 1313 ACATATATGCTCTACTTATATAGAGAGATCTTAAGGTTCATATGTTCAAGA 1362
 367 uasnglyPheLeuSerSerLeuProTyrlleuGlySerTrpLeuGlySmetI 384

1363 GAATGAGTTTATATCTTCATTCAGCTTATTTAGGCTTGTGTTATGTATGA 1412
 384 LeuSerGlyGlnAlaAlaAspAsnLeuArgAlaLysTrpAsnPheSer 400
 1413 TCTCTGTGGTCAAGCTCTGTGACAAATTTTAAGGGCAAAATTTGAAATTTTCA 1462
 401 ThrLeuCysValArgArgIlePheSerLeuIleGlyMetIleGlyProAl 417
 1463 ACTTATATGTGTGCGAGAAATTTTATGCTTATAGAAATGATTTGACCTGC 1512
 417 aValPheLeuValAlaAlaLysPheIleGlyCysAspTySerLeuAlav 434
 1513 AGTATTCCTGTAGTGTGCTGCTTCATTCATGCTGATGATTTCTTTGGCG 1562
 434 aAlaPheLeuThrIleSerThrThrLeuGlyGlyPheCysSerSerGly 450
 1563 TTGCTTTCCTAATATATATCAACACACTGGAGGCTTTTGTCTTCTGGA 1612
 451 PheSerIleAsnHisLeuAspIleAlaProSerTyraLysIleLeuLe 467
 1613 TTTAGCATCAACCATCTGATATTTGCTCTTCGTATGCTGATATCTCTCT 1662
 467 uGlyIleThrAsnThrPheAlaThrIleProGlyMetValGlyProValI 484
 1663 GGGCATCAAAATACATTTGCCACTATTCAGAGATGTTGGGCGCCGTCA 1712
 484 LeAlaLysSerLeuThrProAspAsnThrValGlyGluTrpGlnThrVal 500
 1713 TTGCTTAAAGTGTGACCCCTATATACACTGTGTGGAGATGGCAAAACGTG 1762
 501 PheTyrlleAlaAlaAlaIleAsnValPheGlyAlaIlePhePheThrLe 517
 1763 TTTATATATGCTGCTGCTATTAATGTTTTTGTGTCATTTCTTTTACACT 1812
 517 uPheAlaLysGlyGluValGlnAsnTrpAlaLeuAsnAspHisGlyH 534
 1813 ATTGCGCAAAAGTGAAGTACAAACTGGGCTCAATGATCATCATGATGAC 1862
 534 LsArgHis 536
 1863 ACAGACAC 1870
 seq_name: gb_pr:AF244577
 seq_documentation_block: 2930 bp mRNA linear PRI 06-AUG-2000
 LOCUS AF244577
 DEFINITION Homo sapiens membrane glycoprotein HP59 (HP59) mRNA, complete cds.
 ACCESSION AF244577
 VERSION AF244577.1 GI:9719373
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 2930)
 Fu, C., Barhan, S., Ceteanu, N.D., Lloyd, S.R., Yan, H.-P.,
 Carter, C.E., Shi, E., Venkov, C., Yakes, M.F., Page, D.L. and
 Hellerqvist, C.G.
 Identification of a novel membrane protein from mammalian cells
 that interacts with the anti-pathogenic compound CM101
 JOURNAL
 Unpublished
 2 (bases 1 to 2930)
 Fu, C., Barhan, S., Ceteanu, N.D., Lloyd, S.R. and Hellerqvist, C.G.
 Direct Submission
 Submitted (13-MAR-2000) Biochemistry, Vanderbilt University, School
 of Medicine, 23rdPierce, Nashville, TN 37232-0146, USA
 FEATURES
 Location/Qualifiers
 1..2930
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..2930
 /gene="HP59"
 gene

CDS 263..1873
/gene="HP59"
/note="similar to Na⁺ inorganic phosphate transporter"
/codon_start=1
/product="membrane glycoprotein HP59"
/protein_id="AA097469.1"
/db_xref="GI:9719374"
/translation="MAAGAMPPPRPVOPAPRGFGSLGSRSLTCQVASTPAHVGMS
PVRDARDGEESTDRPIPLPGAPRAEAPVCGSARYNALIAFEFPIYALRVNLS
VALVADMSNTLTEDNRTSKACPEHSAPIRVHNQTKKKYQMDAETGGMLISFPFVY
IITQIPGGYASKIGKMLGFLGTAVALTLFPIAADLGVGLIYRALLEGLEGV
TFPMAHAWSSWAPLESKLISISYGAQLGTVISLPSGICTYMMWTVFFEFST
IGIFWFLIMLVSDTPQKHRISEYKEYLISLSRLQSSQKSVPVVPIILKSLPMA
IVVAHESYNTFFYTLTLPLPYMKEILRFNOENGFLSLPYLGSMLCMILSGOASN
LRAMNFSITLCVRIFSLIGMIPAVFLVAGFVGCQYSLAVAFITLSTLLGFCSSG
FSIHNDIAPSYACILIGITVFATITPGMVPYIAKSLTPONTGEMQVYVYIAALIN
VFQAIFFLFLFKGEVQNMALNDHGHHR"

BASE COUNT 729 a 698 c 681 g 822 t
ORIGIN

alignment_scores:
Quality: 2836.00 Length: 536
Ratio: 5.291 Gaps: 0
Percent similarity: 100.000 Percent identity: 100.000

alignment_block:
US-09-776-865-2 x AF244577 ..

Align seg 1/1 to: AF244577 from: 1 to: 2930

1 MetaLaalaglyAlaMetThrProArgProValGlnProAlaArgpr 17
|||||
263 ATGCGCGCGGGGCGATGACACCGCCGCCGCTCCACGCGACTGGCC 312
17 oglyglypheglyLeuSerCylArgArgSerLeuLeuCysGlnValAlas 34
|||||
313 CGGGGCGCTCGGGCTCGGGCGCGGCGCTCTCTCTGCGAGTGGCGCA 362
34 erThrProAlaHisValGlyValMetArgSerProValArgAspLeuAla 50
|||||
363 GTACACCTGCTACGTAGGCGGTACAGGTCTCCGGTCCAGACCTGGCC 412
51 ArgAsnAspGlyGlnGlnSerThrAspArgThrProLeuLeuProGlyAla 67
|||||
413 CGGAGCGATGGCGAGAGACGACGACGCGCTCTCTTACCGGGCGC 462
67 apProArgAlaGlnAlaAlaProValCysCysSerAlaArgTyraAsnLeu 84
|||||
463 CCCAGCGCGCGAAGCGCGCTCCAGTGTGCTGCTGCTTACCAACTTGA 512
84 laLeuLeuAlaPhePheglyPhePheIleValTyraLeuArgValAsn 100
|||||
513 CAATTTGGCGCTTTTGTGTCTTCATGTGTATCATTAAGTGTGAT 562
101 LeuSerValAlaLeuValAspMetValAspSerAsnThrThrLeuGlnAla 117
|||||
563 CTGAGGTGTCGCTTACTGATAGTAGATTCAAATACAACTTTAAGA 612
117 pasnArgThrSerLeuAlaCysProGlnHisSerAlaProIleLeuValAla 134
|||||
613 TAATAGAACTCCAGAGCGGTCCAGAGCATCTGCTCCCATAAAGTTC 662
134 isHisAsnGlnThrGlyLysLysTyrglnTrpAspAlaGlnThrGlnGly 150
|||||
663 ATCATATACAAAGGAGTACCAAGGATGACAGAACTCAAGA 712
151 TrpIleLeuGlySerPhePheTyrglyTyraIleThrGlnIleProGly 167
|||||
713 TGGATTCGCGTCTTTTATGCGTACATCATCACAGATTCTCTGG 762
167 yglyTyraValAlaSerLysIleGlyLysMetLeuLeuGlyPheGlyAla 184
|||||

763 AGGATATGTTGCCAGCAAAATAGGGGGAATAATGCTGTAGATTTGGGA 812
184 leLeuGlyThrAlaValLeuThrLeuPheThrProIleAlaAlaAspLeu 200
|||||
813 TCCTTGGCACTGCTGCTCCACCCCTTCTCAGTCCCTGCTGAGATTTA 862
201 glyValGlyProLeuIleValLeuArgAlaLeuGlyLysLeuGlyGln 217
|||||
863 GGAGTGGAGCAACCTATGACTACAGAGCACTAAGAAGAGCTAGAGAGG 912
217 yValThrPheProAlaMetHisAlaMetTrpSerSerTrpAlaProPro 234
|||||
913 TGTATCATTTCCAGCCATCATCCATGTCGCTCTTCTGGCGCCCTCC 962
234 engLysArgSerLysLeuLeuSerLysSerTyraAlaGlnIleLeuGly 250
|||||
963 TTGAAGAGCAACACTTCTTAGCATTTGATGACAGAGACAGCTTGGG 1012
251 ThrValIleSerLeuProLeuSerGlyIleIleCysTyraTyraMetAsn 267
|||||
1013 ACAGTAATTTCTCTCTCTCTCTGGAATATTTGCTACTATATGAATG 1062
267 pThrTyraValPheTyraPhePheGlyThrIleGlyIlePheThrPheLeu 284
|||||
1063 GACTATATGCTCTCTACTTTTGTGCTATGGAATATTTGGTCTCTT 1112
284 eutRpiLeuThrLeuValSerAspThrProGlnHisLysArgIleSer 300
|||||
1113 TGTGATCTGCTGATGATGACACACCAAAACACAGAAATTTCC 1162
301 HisTyrglyLysGlnTyraIleLeuSerSerLeuArgAsnGlnLeuSer 317
|||||
1163 CATTATGAAGAAGATATACATTTCTTCATCATTAAGAATACAGCTTCTC 1212
317 rgLysSerValProTrpValProIleLeuLysSerLeuProLeuTrp 334
|||||
1213 ACAGAAAGTACGTCCTGGGTACCCATTTAAATCCCTGCGCACTTGGG 1262
334 laLeuValAlaHisPheSerTyraAsnTrpPheTyraThrLeuLeu 350
|||||
1263 CTATCGTAGTGCACACTTTCTTACACGACCTTTTATATATCTTAATG 1312
351 ThrLeuLeuProThrTyraMetLysGlnIleLeuArgPheAsnValGln 367
|||||
1313 ACATTTATGCTCTATATATGAAGAGATCTTAAGTTCAATGTTAAGA 1362
367 uAsnGlyPheLeuSerSerLeuProTyraLeuGlySerTrpLeuCysMet 384
|||||
1363 GAATGGGTTTATTTCTTCAATGCTTATTTAGGCTCTTGATATGATGA 1412
384 leLeuSerGlyGlnAlaAlaAspAsnLeuArgAlaLysTrpAsnPheSer 400
|||||
1413 TCCTGTGCTGCAAGGCTGACAAATTTAAGGGCAAAATGAATTTTGA 1462
401 ThrLeuCysValArgArgIlePheSerLeuIleGlyMetIleGlyProAla 417
|||||
1463 ACTTTATGCTTGCAGAAATTTTACCTTATGAGATGATGAGACTGCG 1512
417 aValPheLeuValAlaAlaGlyPheIleGlyCysAspTyraSerLeuAla 434
|||||
1513 AGTATTCCTGCTAGCTGCTGCTTCAATGCTGATTTATTTCTTGGCG 1562
434 alaAlaPheLeuThrIleSerThrThrLeuGlyGlyPheCysSerSerGly 450
|||||
1563 TTGCTTTCTTAATATATCAACAACCTGGAGGCTTTGCTCTTCTGGA 1612
451 PheSerIleAsnHisLeuAspIleAlaProSerTyraAlaGlyIleLeu 467
|||||
1613 TTATAGATCAACCATCTGATATGCTCTGCTATGCTGATCTCTCTCT 1662
467 uGlyIleThrAsnThrPheAlaThrIleProGlyMetValGlyProVal 484
|||||
1663 GGGCATCAACAATATATGCTGCTATTCAGGAATGTTGGGCGCGTCA 1712

```

484 1eAlaIysSerLeuThrProAspAsnThrValIGlyLutRPGlnThrVal 500
|||||
1713 TTGCTAAAAGTCGACCCCTGATACACTGTGGAGAAATGCAACCTG 1762
501 PheTyrIleAlaAlaIleAsnValPheGlyAlaIlePhePheThrLe 517
|||||
1763 TTCATATATGCTGCTGCTATTAATGTTTGTGTCATTTCTTTACACT 1812
517 uPheAlaIysGlyIGlyValGlnAsnTPalaLeuAsnAspHisHSGlyH 534
|||||
1813 ATTCGCCAAAGTGAAGTACAAACTGGCTCCTCAATGATCACCATGGAC 1862
534 ISARGHIS 536
|||||
1863 ACAGACAC 1870

seq_name: gb_pat:AX138494

seq_documentation_block:
LOCUS AX138494 2512 bp DNA linear PAT 30-MAY-2001
DEFINITION Sequence 2 from Patent EP1069104.
ACCESSION AX138494
VERSION AX138494.1 GI:14274389
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS
TITLE Human anion transporter gene implicated in salt disease and
lysosomal static acid transport
JOURNAL Patent: EP 1069104-A 2 17-JUN-2001;
Akzo Nobel N.V. (NL)
FEATURES
source location/Qualifiers
1..2512
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 612 a 585 c 566 g 749 t
ORIGIN

alignment_scores:
Quality: 2804.00 Length: 535
Ratio: 5.261 Gaps: 1
Percent Similarity: 99.626 Percent Identity: 99.252

alignment_block:
US-09-776-865-2 x AX138494
Align seg 1/1 to: AX138494 from: 1 to: 2512

```

```

85 1eAlaIaphPheGlyPhePheIleValTyrAlaLeuArgValAsnLeu 101
|||||
401 TTGTGGCTTTTGTGTTCTTCATTTGTATGATTCATACGTGAATCTG 450
102 SerValAlaLeuValAspMetValAspSerAsnThrThrLeuGlnAsp 118
|||||
451 AGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 500
118 nArgThrSerLysAlaCysProGlnHisSerAlaProIleLysValHis 135
|||||
501 TAGAAGCTTCAGAGCGGTGTCAGAGCATTCGCTCCCATTAAGTTATC 550
135 ISANGINTHRGLYLYSLYSTYRGINTRPASPALAGTURGINGLYTRP 151
|||||
551 ATATCAACGGGTATAGAGATGACCAATGGGATCGAAGAACATCAAGATG 600
152 IleLeuGlySerPhePheTyrGlyTyrIleIleThrGlnIleProGly 168
|||||
601 ATCTCGGTTCTCTTTTATAGCTACATCATCACACAGATTCCTGGAGG 650
168 YTYRValAlaSerLysIleGlyGlyLysMetLeuLeuGlyPheGlyIle 185
|||||
651 ATATGTTGCCAGCAAAATAGGGGGGAAATGCTCTAGGATTTGGATTC 700
185 euGlyThrAlaValLeuThrLeuPheThrProIleAlaAlaAspLeuGly 201
|||||
701 TTGGCACTGCTGCTCCTCACCTGTCACCTCCATGCTGCGAGATTAGGA 750
202 ValGlyProLeuIleValLeuArgAlaLeuGlnGlyLeuGlyGlyVal 218
|||||
751 GTTGACACCACTCATTTACTACAGACACTAGAAAGACTAGACAGCGGTG 800
218 lThrPheProAlaMetHisAlaMetTrpSerSerTrpAlaProLeuG 235
|||||
801 TACATTTCCAGCCATGCAATGCAATGCTCTCTGGGCTCCCTCTTG 850
235 lArgSerLysLeuLeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThr 251
|||||
851 AAAGAACCAAACTTCTTAGCATTTTCATATGAGAGACAGAGCTGGACA 900
252 ValIleSerLeuProLeuSerGlyIleIleCysTyrTyrMetAsnTrp 268
|||||
901 GTAATTTCTCTCTCTCTTCCGAAATATTTGCTACTATATGAAATGGAC 950
268 rTyrValPheTyrPhePheGlyThrIleGlyIlePheTrpPheLeu 285
|||||
951 TTATGCTCTACTTTTGTGTAATGGAATATTTGGTTCTTTGT 1000
285 rPileTrpLeuValSerAspThrProGlnLysHisLysArgIleSerHis 301
|||||
1001 GGATCTGTTATGATGACACACACACAAACCAAGAGATTTCCCAT 1050
302 TyrGlnLysGlnTyrIleLeuSerSerLeuArgAsnGlnLeuSerSer 318
|||||
1051 TATGAAGAAGAAATCATTTCTTCATTAAGAAATACAGCTTCCTGCACA 1100
318 nLysSerValProTrpValProIleLeuLysSerLeuProLeuTrpAla 335
|||||
1101 GAAATCAGTGGCGGTGATCCCATTTTAAATCCCTCCACTTGGGCTA 1150
335 lValValAlaHisPheSerTyrAsnTrpIlePheTyrThrLeuLeuThr 351
|||||
1151 TCGATGTGCACACTTTTCTTACACTGGACTTTTATCTTATTTAGACA 1200
352 LeuLeuProThrTyrMetLysGlnIleLeuArgPheAsnValGlnLys 368
|||||
1201 TTATGCTCTACTTATATGAAGAGATCTTAAGGTTCATATGTTCAAGAGA 1250
368 nGlyPheLeuSerSerLeuProTyrLeuGlySerTrpLeuCysMetIle 385
|||||
1251 TGGGTTTTTATCTTCATTTGCTTATTTAGGCTCTTGGTTATGATGATCC 1300
385 euSerGlyGlnAlaIleAspAsnLeuArgAlaLysTrpAsnPheSerThr 401

```

```
|||||
1301 TGTCTGCTCAAGCTGCTGACATTTAAGGGCAAAATGCAATTTTCAACT 1350
402 LeuGysValArgArgIlePheSerLeuIleGlyMetIleGlyProAlaVala 418
1351 TTATGTGTTCCGAGAAATTTTACCTTAAGGAATGATTGGACCTGCAGT 1400
418 IPhelLeuValAlaIleGlyPheIleGlyCysAspIleSerLeuAlaVala 435
1401 ATTCTGTGAGCTGCTGCTTTCATTTGGCTGTATTTCTTTGGCCGTTG 1450
435 IAPheLeuThrIleSerThrThrIleuGlyGlyPheCysSerSerGlyPhe 451
1451 CTTTCTTAACGTATATCAACAACACGTGGAGCTTTTGTCTTCTGCAATTT 1500
452 SerIleAsnIleLeuAspIleAlaProSerGlyAlaGlyIleLeuLeuGI 468
1501 AGCATCAACCACTGTGATATGCTCTTCGTATGCTGTATGCTCTCTGGG 1550
468 YIleThrAsnThrPheAlaThrIleProGlyMetValGlyProAlaIleA 485
1551 CATCAACAATACATTTGCCACATATCCAGAAATGTTGGCCCGTCATTTG 1600
485 IAlaYSerLeuThrProAspAsnThrValGlyGlyIleProGlnThrValPhe 501
1601 CTAAGAAGCTGACCCCTGATACACTGTTCGAGAAATGGCAACCGTGTTC 1650
502 TTTTleAlaIleAlaIleAsnValPheGlyAlaIlePhePheThrLeuPhe 518
1651 TATATGTCGCTGCTATTAATGTTTGGTGCATTTCTTTACACTAAT 1700
518 eAlaIaYSerGlyIleValGlnAsnThrPAlaLeuAsnAspHisIleGlyHisA 535
1701 CCCCAGAGGTGAAGTACAAACTGGGCTCTCAATGATCAATCAGTGAACACA 1750
535 rGHis 536
1751 GACAC 1755
seq_name: gb_HSA387747
LOCUS HSA387747 2512 bp mRNA linear PRI 10-DEC-1999
DEFINITION Homo sapiens mRNA for sialin.
ACCESSION AJ387747
VERSION AJ387747.1 GI:6562532
KEYWORDS sialin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Verheijen, F.W., Verbeek, E., Aula, N., Beeren, C.E., Havelaar, A.C.,
Jooose, M., Peltonen, L., Aula, P., Galjaard, H., van der Spek, P.J. and
Mancini, G.M.
TITLE A new gene, encoding an anion transporter, is mutated in sialic
acid storage diseases
JOURNAL Nat. Genet. 23 (4), 462-465 (1999)
MEDLINE 20047778
REFERENCE
AUTHORS Verheijen, F.W.
TITLE Direct Submission
JOURNAL Submitted (03-JUN-1999) F.W. Verheijen, Erasmus University, Dept.
Clinical Genetics, P.O. Box 1738, 3000 DR Rotterdam, NETHERLANDS
FEATURES
source
1..2512
/organism="Homo sapiens"
/db_xref="taxon:9606"
271..1758
/codon_start=1
/product="sialin"
/protein_id="CA862540.1"
/db_xref="GI:6562532"
```

```
/translation="MRSPVRLARNDEESTDRTPLPGAPRAEAAPVCCSARYNLAI
LAFGEFFIYVALRVNLISVALVDWDSVTLIEDNRTSACEBAPIKVHNQKXKO
WDAEFGMILGSPFFGYITPOLRGYVASKIGKMLGREGILGTAVILFTPTIADIG
VGPLYLIRALEGIEGEGYTPPAMAMMSNAPPLERKSILSISAGQOLGTVISLPSG
ITCYVMWTVYVFFGTIGLFWFLMWLWVSDTPQKHRIshYEKEYITLSSLNQLSS
OKSPVWPILKSLPLMAIVVAHPSYWNTEFTYLLPTVAKELIRNVQENGLSSIP
YLSMELCMILSGOADNLRANKNFSVLCVARISFLGMLGPVAVLVAAGTIGDYSIA
VAFLLSTLTGFCSSGFSINHLDIAPSVAGIILGTTNFPATIPGAVPIAKSLPDP
NTVEMQTVYIIAAINVEGATFEFTLEFACGEVNMALNDHGRH"
```

BASE COUNT 612 a 585 c 566 g 749 t

ORIGIN

alignment_scores:

| Quality: | 2804.00 | Length: | 535 |
|---------------------|---------|-------------------|--------|
| Ratio: | 5.261 | Gaps: | 1 |
| Percent Similarity: | 99.626 | Percent Identity: | 99.252 |

alignment_block:

us-09-776-865-2 x HSA387747 ..

Align seg 1/1 to: HSA387747 from: 1 to: 2512

```
4 GlyAlaMetThr.....ProProArqProValGlnProAlaArgProGI 18
|||||
151 GGGGGGGGTGTCATCCGCCCGCCCGCGGTCAGCCAGCTCGGCCGG 200
18 yGlyPheGlyLeuSerGlyAlaArgSerLeuLeuGlyValAlaSerT 35
|||||
201 GGGCTTGGGGGTGGGGCGGGCGCTCCCTCTCTGCGCAGGCGGCGAGTA 250
35 hrProAlHHisValGlyAlaMetArgSerProValArgAspLeuAlaArg 51
|||||
251 CACCTGCTCAGCTAGCGCTCATGAGAGTCTCCGTTTCGAGACCTGGCCGG 300
52 AsnAspGlyGlnGluSerThrAspArgThrProLeuLeuProGlyAlaPr 68
|||||
301 AACGATGGCGAGAGAGACGACGACCGCAGCCCTCTTACCGGGCGCCCG 350
68 cArgAlaGluAlaAlaProValCysCysSerAlaArgTyrAsnLeuAlaI 85
|||||
351 ACGGGCCGAGAGCGCTCCAGTGTCTGCTGTCTGCTGTACCAACTTAGCA 400
85 IeLeuAlAPhePheGlyPhePheIleValITyrAlaLeuArgValaAsnLeu 101
|||||
401 TTTTGCCCTTTTGTGTTCTTTCATTTGTGTATGCTATACGTGTGAATCTG 450
102 SerValAlaLeuValaIAspMetValaIAspSerAsnThrThrLeuGluAspAs 118
|||||
451 AGTGTTCGCTTAGTGAATATGATTAATCAATCAACTTATGAGAGATTA 500
118 nArgThrSerIysAlaCysProGlnHisSerAlaProIleGlyValHisH 135
501 TAGAAGCTTCCAGAGCGGTGCCAGACCATTCGCTCCATAAAGTTCATC 550
135 ISAsnGlnThrGlyIlySlySTyrGlnTPAspAlaGlnThrGlnGlyTyr 151
551 ATAAATCAACGGGTAAAGATACCAATGGATGCGAAGAACTCAAGATGAG 600
152 ILeuGlySerPhePheTyrGlyTyrIleIleThrGlnIleProGlyGI 168
601 ATTCTGCGTCTCTTTTATGCGTACATCAACAGATTCCTCGAGAG 650
168 YTyrValAlaSerIysIleGlyGlyMetLeuLeuLeuGlyPheGlyIleL 185
651 ATATGTTCACAGCAAAATAGGGGGAATGCTGTAGATTTTGGATTC 700
185 euGIYThAlaValLeuThrLeuPheThrProIleAlaAlaAspLeuGI 201
701 TTGGACACTGCTGCTCCTCAGCTGTTCATCCATTTGCGAGATTTAGGA 750
202 ValGlyProLeuIleValLeuArgAlaLeuGlnGlyLeuGlyGlnGlyIyVa 218
|||||
```

751 GTTGGACCACTGATTGATCTAGACAGACATGAGAGACTAGAGAGGGTGT 800
 1TrhPheProAlaMethAlaMetTrpSerSerTrpAlaProProLeug 235
 801 TACATTTCCAGCCATGCATGCTGCTCTTGGGCTCCCTCTTG 850
 235 LuArSerLysLeuLeuSerLLeSerTyraLagLalaGlnLeuGlyThr 251
 851 AAAGAACCAAACTTTCATGATTCATGATGAGAGACAGACTGGGACA 900
 252 VallSerLeuProLeuSerGlyLLeLeCysTyrrTyrrMetAnTrpTh 268
 901 GTAATTTCTCTCTCTTCGAAATATTTGCTACATATATGAAATTTGAC 950
 268 rTyValPheTyrrPheGlyThrLLeGlyLLePheTrpPheLeuLeut 285
 951 TTATGCTTCTCACTTTTGGTACTATTTGGAATTTTGGTTCTTTTGT 1000
 285 rPlLeTrpLeuValSerAspThrProGlnLysHisLysArgLLeSerHis 301
 1001 GGATCTGTTAGTAGTACACACACAAACACAGAGAAATTTCCCAT 1050
 302 TyrrLLeuLysGluTyrrLLeLeuSerSerLeuArGAsnGlnLeuSerG 318
 1051 TATGAAAGGAATACATTTCTTCATTCATTAAGAAATCAGCTTCTTACA 1100
 318 nLysSerValProTrpValProLLeuLysSerLeuProLeuTrpAla 335
 1101 GAATCTAGTGGCGGTGGGTACCATTTTAAATCCCTCCACTTTGGGCTA 1150
 335 LeValValAlaHisPheSerTyrrAsnTrpThrPheTyrrThrLeuLeuThr 351
 1151 TCGAGTAGTCACACTTTTCTTACAACTGAGCTTTTATCTTATGTGACA 1200
 352 LeuLeuProThrTyrrMetLysGlnLLeLeuArGpHeAsnValGlnLLeu 368
 1201 TTATGTGCTACTTATATGAGGAGATCTTAAAGTTCAATGTTCAAGGAA 1250
 368 nGlyPheLeuSerSerLeuProTyrrLeuGlySerTrpLeuCysMetLLe 385
 1251 TGGGTTTTTATCTTCACTTGGCTTATTTAGGCTCTTGGTTATGATGATCC 1300
 385 eUsErGlyGlnAlaAlaAspAsnLeuArGAlaLysTrpAsnPheSerThr 401
 1301 TGCTGTGCAAGCTGCTGACAAATTTAAGGCAAAATTTTTCAACT 1350
 402 LeuCysValArGArgLLePheSerLeuLLeGlyMetLLeGlyProAlaVa 418
 1351 TTATGTGTGCGAGAAATTTTATGACCTTATGGAATGATGACCTGCAGT 1400
 418 lPheLeuValAlaAlaGlyPheLLeGlyCysAspTyrrSerLeuAlaVala 435
 1401 ATTCTGTGAGCTGCTGCTTCATTTGCTGATTTATCTTTGGCCGTG 1450
 435 lAphLeuThrLLeSerThrThrLeuGlyGlyPheCysSerSerGlyPhe 451
 1451 CTTTCTCAACTATATCAACAACACTGGAGGCTTTTCTCTTCGAGATT 1500
 452 SerLLeAsnHisLeuAspLLeAlaProSerTyrrAlaGlyLLeLeuLeuG 468
 1501 AGCATTCACACATCGGATATGCTCTCTTCATGCTGATGATCCCTCCGG 1550
 468 yLLeThrAsnThrPheAlaThrLLeProGlyMetValGlyProValLLea 485
 1551 CATTACAAATACATTTGCCACTATTCAGAGAAATGTTGGGCCCTGCTAT 1600
 485 lAlysSerLeuThrProAspAsnThrValGlyLLeuTrpGlnThrValPhe 501
 1601 CTAAGAACTGAGCCCGATACACACTGTGAGAAATGCAAAACGCTGTC 1650
 502 TyrrLLeAlaAlaAlaLeuAsnValPheGlyAlaLLePhePheThrLeuPh 518
 1651 TATATGCTCTGCTATTAATGTTTTTGGTCCATTTTCTTACACATATT 1700

518 eAlaLysGlyLLeuValGlnAsnTrpAlaLeuAsnAspHisGlyLLeHis 535
 1701 CGCCAAAGGTGAGAGTACAAACTGGGCTCTCATGATGATCACCATGAGACACA 1750
 535 rGHis 536
 1751 GACAC 1755
 seq_name: gb_pr:BC020961
 seq_documentation_block:
 LOCUS BC020961 3329 bp mRNA linear PRI 22-JAN-2002
 DEFINITION Homo sapiens, solute carrier family 17 (anion/sugar transporter), member 5, clone MGC:8885 IMAGE:3847279, mRNA, complete cds.
 ACCESSION BC020961
 VERSION BC020961.1 GI:18088702
 KEYWORDS MGC.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 3329)
 Strausberg, R.
 Direct Submission
 Submitted (03-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) md@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 20 Row: 1 Column: 2
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6912665.
 FEATURES
 source
 Location/Qualifiers
 1..3329
 /organism="Homo sapiens"
 /db_xref="locusID:26503"
 /db_xref="taxon:9606"
 /clone="MGC:8885 IMAGE:3847279"
 /tissue_type="Colon, adenocarcinoma"
 /clone_id="NIH_MGC_65"
 /lab_host="DH10B"
 /note="Vector: pcMV-SPORT6"
 125..1612
 /codon_start=1
 /product="solute carrier family 17 (anion/sugar transporter), member 5"
 /protein_id="AAH20961.1"
 /db_xref="gi:18088703"
 /translation="MRSVRDLARNDSGESEDRPRLPGARAAAPVCCSARYNLILAFGEPTIVALLRNLSALVDWDNSNTLEDNRTSRACPEHSAPPIVHHNQTKKYWDAEFGWILDSFFYGIITQIPGIVASKIGKMLGFGILGVALVLPPLIADYG VGPLIVLALGLEGVTFPMAHMAWSSWAPLERSKILSISYGAOLGTVLSLPLSG IICYYMNTYVFEFGIGIFWFLMLIVLSDFPQKRISHYEKEYILSLRLSS OKSDWVPIELSLPLMAIVVAHFSYNTMFYLLLPYKMEILRFNVNVEGFLPSIAYGSMICWILSGAANDLRAKWNSFTLCVRPISLIMIGGPAVLVAAGTFCYSILA VAPLITSTLOGRCCSSGSINHDIDIASYKAILLGITNTPTITFGMGVPIYAKSLTPD NTVGEWCVTFEITAAINVEGAIFFTLFRKGEVQWMLNDHRGHH"
 BASE COUNT 863 a 718 c 727 g 1021 t

ORIGIN

alignment_scores:

Quality: 2770.50 Length: 528
 Ratio: 5.257 Gaps: 1
 Percent Similarity: 99.811 Percent Identity: 99.432

alignment_block:

us-09-776-865-2 x BC020961 ..

Align seg 1/1 to: BC020961 from: 1 to: 3329

```

10 ArgProvalGlnPro...AlaArgProGlyGlyPheGlyLeuSerGlyIar 25
||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
26 CGACCCAGCGGTCGGGCGGCTCGGCGGCGGCTTCGGGCTGTGGGCGG 75
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
25 GARSerLeuLeuGlyGlnValAlaSerThrProAlaHisValGlyAlaM 42
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
76 GCGCTCCCTCTCTGCGCAGGTGGCGAGTACACTGCTCAGCTAGGCGTCA 125
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
42 eTArgSerProValArgAspLeuAlaArgAsnAspGlyGluSerThr 58
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
126 TGAAGTCTCGGTTGAGACCTGGCCGGAACGATGGCGAGAGACAGCAG 175
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
59 AsparGThrProLeuLeuProGlyAlaProArgAlaGluAlaAlaProVa 75
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
176 GACGCGACGCGCTCTTCTACGCGGCGGCCCGACGCGGCGGCTCCAGT 225
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
75 lCysCysSerAlaArgTyrAsnLeuAlaIleLeuAlaPhePheGlyPhe 92
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
226 GTGCTGCTGCTGCTGCTACCACTTACCAATTTTGGCTTTTGGTTTCT 275
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
92 heIleValTyrAlaLeuArgValAsnLeuSerValAlaLeuValAspMet 108
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
276 TCATGTGTGATGCAATGATGATGATGATGATGATGATGATGATGATG 325
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
109 ValAspSerAsnThrThrLeuGluAspAsnArgThrSerLysAlaCysPr 125
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
326 GTAGATTCAATAATACACTTTAGAAATATATGAACCTTCGAAGCATGACC 375
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
125 OGluHisSerAlaProIleLeuValHisHisAsnGlnThrGlyLysLysT 142
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
376 AGACCATTCCTGCTCCATAAAGTTCAATCAATCAAAACGGGTAAAGAT 425
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
142 YrGlnTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTyr 158
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
426 ACCAATGGATGCAAGAACTCAAGATGATGATGATGATGATGATGATGAT 475
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
159 GlnTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerLysIleG 175
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
476 GCGTACATCATCACACAGATTCCTGGAAGATATGTTGCCACCAAAATAGG 525
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
175 yGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaValLeuThrL 192
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
526 GGGGAAATGCTGCTAGGATTTGGGATCTTGGCATGCTGCTCTCACC 575
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
192 eupheThrProIleAlaAlaAspLeuGlyValGlyProLeuIleValLeu 208
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
576 TGTTCACCTCCCATGCTGCAAGATTAGAGTTGACCACTCATTTGTACTC 625
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
209 ArgAlaLeuGlnGlyLeuGlyGluGlyValThrPheProAlaMetHisAl 225
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
626 AGACCACTAGAGAGCTAGAGAGAGGTGTATCATTTCCAGCCATGCAATGC 675
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
225 aMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuSerI 242
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
676 CATTTGGTCTCTTGGGCTCCCTCTTGAAGAAGCAAACTTCTTAGCA 725
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
242 leSerTyrAlaGlyAlaGlnLeuGlyThrValIleSerLeuProLeuSer 258
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
726 TTTCAATGACGAGACACAGCTTGGACAGTAATTTCTTCTTCTCTTCT 775

```

```

259 GlyIleIleCysTyrTyrMetAsnTrpThrTyrValPheThrPhePheG 275
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
776 GGAATTAATTTGCTACTATATGATTAATGAGCTTATGCTTCTTCTTTGG 825
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
275 yThrIleGlyIlePheThrPheLeuLeuTrpIleThrLeuValSerAsp 292
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
826 TACTATTTGGAATTAATTTGTTGTTCTTTGTGATCTGTTAGTAGTGACA 875
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
292 hrProGlnLysHisLysArgIleSerHisTyrGlnLysGluTyrIleLeu 308
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
876 CACCACAAAACACACAGAGAAATTTCCATTATGAAAAGAAATACATCTCT 925
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
309 SerSerLeuArgAsnGlnLeuSerSerGlnLysSerValProTrpValPr 325
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
926 TCATCATTTAAGAATTCAGCTTCTTTCACAGAAAGCAGCGCGCTGGATCC 975
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
325 oIleLeuLysSerLeuProLeuTrpAlaIleValAlaAlaHisPheSer 342
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
976 CATTTTAAATCCCTGCCACTTGTGGCTATCGTAGTTCACACCTTTCTT 1025
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
342 yTrsTrpThrPheTyrThrLeuThrLeuThrLeuProThrTyrMetLys 358
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1026 ACAACCTGACTTTTATTACTTATTTGACATTATTTGCCCTATTATGAA 1075
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
359 GluIleLeuArgPheAsnValGlnGluAsnGlyPheLeuSerSerLeuPr 375
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1076 GAGATCCCTAAGTTCAMTGTTCACAGACAGATGCGTTTATCTTATGCC 1125
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
375 oTyrLeuGlySerTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 392
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1126 TTATTTAGGCTCTGTTATGATGATCCTGCTGCTCAAGCTCTGACA 1175
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
392 snLeuArgAlaLysTrpAsnPheSerThrLeuCysValArgIleIlePhe 408
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1176 ATTTAAGGCAAAATGGAATTTTCAACTTATATGTGTGGAGAAATTTT 1225
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
409 SerLeuIleGlyMetIleGlyProAlaValPheLeuValAlaAlaGlyP 425
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1226 AGCTTATAGGAATGATTTGACCTGCACTGATTTCTGGTAGCTGTGGCT 1275
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
425 eIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 442
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1276 CATTTGGCTGTGATTAATCTTTGGCGGCTTCTTCTCACTATATCAACA 1325
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
442 hrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIle 458
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1326 CACTGGAGGCTTTTGCTCTTCTGATTTACATCAACCACTGGAATATT 1375
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
459 AlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaTh 475
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1376 GCTCCTTCGTATGCTGGATCTCTGCGCATTCACAAATATATTGGCAC 1425
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
475 rIleProGlyMetValGlyProValIleAlaLysSerLeuThrProAspA 492
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1426 TATTTCCAGGAATGTTGGCCCGCTCATTTGCTTAAAGTCTGACCCCTGATA 1475
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
492 snThrValGlyGluTrpGlnThrValPheTyrIleAlaIleAlaIleAsn 508
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1476 ACACCTGTGGAGAAATGGCAACCGTGTTCATATTTGCTGCTGCTATTAT 1525
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
509 ValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnAs 525
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1526 GTTTTGGTGGCATTTTCTTACACTATTTGCCCAAAAGTGAGTAAACAAA 1575
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
525 nTrpAlaLeuAsnAspHisHisGlyHisArgHis 536
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1576 CTGGGCTCTCAATGATCACCATGACACAGACAC 1609
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq_name: gb_cm:AF244578

seq_documentation_block:

LOCUS AF244578 2844 bp mRNA linear MAM 06-AUG-2000
 DEFINITION Ovis aries membrane glycoprotein SP55 (sp55) mRNA, complete cds.


```

1178 AGTTTATAGCTTATAGGATGATGACCTGCATATCTCGTGCCG 1227
423 laaglypheleleaglycysaspyrserleuAlaValAlaphlethrle 439
1228 CAGATTATAGGCTGATTAATTCCTGGCTGTGCAATTCCTAACATA 1277
440 SerThrThrleuaglyglyPheCysSerSerGlyPheSerIleAsnHisle 456
1278 TCACAAACCCGGAGGCTTTTGTCTTCTTGATTTAGCATCAACCATCT 1327
456 uaspllealaproserlyralaglyleleuLeuGlyIleThrAsnThr 473
1328 GGACATGCTCTGCTGATGCTGATTCCTGGGCAATCACAATACCT 1377
473 heaIaThrIleProGlyMetValGlyProValIleAlaIalysSerleuThr 489
1378 TTGCCACTATTCCTGATGATGATGGCCCAATTCGCCAAGCTTACC 1427
490 ProAspAsnThrValGlyGluTrpGlnThrValPheTyrIleAlaIal 506
1428 CCTGAGAACACTATTTGGAGATGCAAACTGTTTGCATCGCTGCTGC 1477
506 alIasnValPheGlyAlaIalIephetherThrLeuPheAlaIalysGlyLuv 523
1478 TATCAATGATTTGCTGCAATTTCTTCACACATTCGCCAAGGTGAG 1527
523 alGlnAsnTrpAlaLeuAsnAspHisHisGlyHisArgHis 536
1528 TGCAAACTGGGCCATCAGTATCACCAGACACAGAAC 1568
seq_name: gb_pat:AX207626
seq documentation block:
LOCUS AX207626 2844 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 3 from Patent WO0156598.
ACCESSION AX207626
VERSION AX207626.1 GI:15422331
KEYWORDS
SOURCE
ORGANISM Ovis sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Caprinae; Ovis.
REFERENCE
1 (bases 1 to 2844)
Hellerqvist, C.G.
METHODS for preventing or attenuating pathoangiogenic conditions by
using the gbs-toxin (cm101) receptor as a vaccine
Patent: WO 0156598-A 3 09-AUG-2001;
JOURNAL VANDERBILT UNIVERSITY (US)
FEATURES
source
1..2844
/organism="Ovis sp."
/db_xref="taxon:9939"
84..1571
/feature="unnamed protein product"
/codon_start=1
/feature="unnamed protein product"
/db_xref="GI:15422332"
/translacion="MKSPPVSDIAPSDGERSDRTPLLORAPAEAPVCCSARYNIAF
LSFPGFVLYSLRVNLSVALVMDVSNNTAKDNRTSYBEAHSAPIKVLAHNTGKKYR
MDAENGMILGSEFGYILTOJPGGVASRGSKLLGIFATATFTPTPLADFG
VGAIVLALBELGEGVTYRPAHAMMSWAPPLERSKLISYAGAGCTGVSLPSG
VICYVYNNVYVYFREGIVGILNFIIMICVSDTPETRIKITYREKYLISLKNQISS
QKSVWIPMLKSLPLMAIVAHFSTWTFYTLTLPLTYMKEVLRNIDENGFLSAVP
YLGCWLCMLISQADNLRAWNEFTLWVRVFSLIGMIGPAIFLVAAGFIQCDVSLA
VAFIISTLGGFCSSGFSINHLDIASVAGILGITTNPATIPGMIGPIIARSILPE
NTIGEMOTVFCIAAIVNFGALIFETLFAKGEVQVMAISDHGHRN"
BASE COUNT 745 a 623 c 594 g 882 t
ORIGIN

```

alignment_scores:

```

Quality: 2338.00 Length: 497
Ratio: 4.901 Gaps: 0
Percent Similarity: 95.976 Percent Identity: 86.117
alignment_block:
US-09-776-865-2 x AX207626
Align seg 1/1 to: AX207626 from: 1 to: 2844
40 GlyValMetArgSerProValAlaArgAspLeuAlaArgAsnAspGlyGlu 56
78 GGCATCATGAAGTCCCGGCTTCGACTTACCCCGAGCGGAGGAGGA 127
56 uSerThrAspArgThrProLeuLeuProGlyAlaProArgIalagluAla 73
128 GGGCTCGGACCGCACACCGCTCTGACAGCGGCCCGGCGGAGACCCG 177
73 laProValCysCysSerAlaArgTyrAsnLeuAlaIleLeuAlaPhePhe 89
178 CTCACGATAGCTGCTGCTGCTGCTTACAACTTACGATTTTGTCTTTT 227
90 GlyPhePheIleValIlyAlaLeuArgValAsnLeuSerValAlaLeuVa 106
228 GGTTCCTTCGTTCTTCTATTCATTCAGGCTGATGCTGACGCTGAC 277
106 laSpMetValAspSerAsnThrThrLeuGluAspAsnArgThrSerLysa 123
278 GGACATGGTGGATTCMAACACACTGCCAAGATATATGAACGCTTACG 327
123 laCysProGlnHisSerAlaProIleLysValHisHisAsnGlnThrGly 139
328 AGTGTGACAGACATTCGCTCCCTCAAAAGTTTTCACAAACCAACGGGT 377
140 LysIstYrGlnTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPh 156
378 AAAAGATACCGGTGGATGACGAAACTCAAGAGATGATTCGAGATCTTT 427
156 ePheTyrGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerL 173
428 TTTCTATGCTACATCATCACACAAATTCGAGAGATTTGCCAGCA 477
173 yslleGlyGlyLysMetLeuLeuGlyPheGlyIleLeuGlyThrAlaVal 189
478 GAATGGGGGAGACTGTTGCTAGGATTCGGAGTCTTGTACACGCTATC 527
190 LeuThrLeuPheThrProIleAlaAlaAspLeuGlyValGlyProLeuII 206
528 TTCACCCCTGTCACCTCCCTGCTGACAGATTCGAGTGGAGCCCTTGT 577
206 eValLeuArgAlaLeuGlyGlyLeuGlyGlyGlyValIValThrProAlaM 223
578 TGCACCTCAGGCACTAGAGGCTAGAGAGGCTGTCACATATCCAGCA 627
223 ethIsAlaMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeu 239
628 TGCATGCCATGTGCTTCATAGGCTCCCTCTTGAAGAAGAGAGCTT 677
240 LeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrValIleSerLeuP 256
678 CTGAGTATTCATATGACAGACAACTTGGACAGAGTTCCTTCC 727
256 oLeuSerGlyIleIleCysTyrTyrMetAsnTrpThrTyrValPheTyrP 273
728 TCTTTCGAGTAAATTTGCTACTATATGAATTTGACCTTATCTCTTAT 777
273 hePheGlyThrIleGlyIlePheThrPheLeuLeuTrpIleTrpLeuVal 289
778 TCTTTGGATGTTGGAATCATCTGTTTATTTATGATCTGCTTACTT 827
290 SerAspThrProGlnLysHisLysArgIleSerHisTyrGlyLysGlyT 306
828 AGTATACACCAAGAACATCACAGACATCTACCTCATGAAAAGAGACTA 877

```

```

306 rTleuSerSerleuArgasnGlnleuSerSerGlnlySerValProT 323
|||||
878 TATCTTCATCATTAATAAAATCAGCTCTCTCACAGAAAGTAGCGCGT 927
|||||
323 rPValProIleuLySerleuProleuTPAlIleValAlaIahIs 339
|||||
928 GGATACCATGCTGAATCACACCTTTGGGCTATTGTCCTTCACAT 977
|||||
340 PheSerTyraSTrPThrPheTrThrleuLeuThrleuProThTy 356
|||||
978 TTTCTTCACTAGGACCTTTTATCTTTGTTGACCTTATTCCTACTTA 1027
|||||
356 rMelysGluIleuArgPheasnValGlnGlnasnGlyPheleuSerS 373
|||||
1028 CATGAAGAAGCTTAAGTTCAATATTCACAGAAAGGTTTATCTCG 1077
|||||
373 eTleuProTyIleuGlySerTrPleuCysMetIleuSerGlyAla 389
|||||
1078 CAGTCCCTTATTAGTTGGTTGTTATGTATCATCTCTCGGTCACCT 1127
|||||
390 AlaasPasnleuArgAlaIySTrPasnPheSerThrleuCysValArg 406
|||||
1128 GCTGCATTTTAGGCAAGATGAAATTTTCACTCTGTGGGTCGAAG 1177
|||||
406 gTlePheSerleuIleGlyMetIleGlyProAlaValPheleuVala 423
|||||
1178 AGTTTATAGCCTTATAGGATGATGACCTGCATATTCCTGTTGCCG 1227
|||||
423 lAglyPheIleGlyCysAspTySerleuAlaValAlaPheleuThr 439
|||||
1228 CAGGATTTATAGCTGTGATTTATCTTGGCTGTGCAATCCTAACATA 1277
|||||
440 SerThrThrleuGlyGlyPheCysSerSerGlyPheSerIleasnIle 456
|||||
1278 TCACAACACCTTGAGGCTTTTGCTCTTGATTTAGCATCAACCATCT 1327
|||||
456 uAspIleAlaProSerTyraIaglyIleleuLeuGlyIleThrAsnTh 473
|||||
1328 GCACATTCCTCTCGTATGCTGATTTCTCTGGGCATCACAAAATACCT 1377
|||||
473 heAlaThrIleProGlyMetValGlyProValIleAlaLySerleuThr 489
|||||
1378 TTGCCATATTCCTGGAATGATGGGCCATCATTCGACAGAGCTTTACC 1427
|||||
490 ProAspAsnThrValGlyIuTrpGlnThrValPheTyIleAlaAla 506
|||||
1428 CCGTGAACACATTTGAGAAAGCAAACTGTTTCTGCATCGCTGCGC 1477
|||||
506 aIleasnValPheGlyAlaIlePhePheThrleuPheAlaLySGlyIuV 523
|||||
1478 TATCAATGATTTGGTGGCATTTTCTTCACACTATTCGCCAAAGGTGAAG 1527
|||||
523 aGlnAsnTrpAlaIleuasnAspHisGlyHisArgHis 536
|||||
1528 TGCAAACTGGGCGCATGATCACCAAGACACAGAAAC 1568
|||||

```

seq_name: gb_in:AY060776

seq_documentation_block:

LOCUS AY060776 1841 bp mRNA linear INV 08-NOV-2001

DEFINITION Drosophila melanogaster GH23975 full length cDNA.

ACCESSION AY060776

VERSION AY060776.1 GI:16768209

KEYWORDS FLI CDNA.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1841)

Stapleton,M., Broksstein,P., Hong,L., Agbayant,A., Carlson,J.,

Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R.,

Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,

NUMCO,J., Pacle,b,J., Paragas,V., Park,S., Phouanavong,S., Wan,K.,
Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.
Direct Submission
Submitted (30-OCT-2001) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

COMMENT

This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our web site
(http://fruitfly.berkeley.edu) or send email to
cnaef@fruitfly.berkeley.edu.

FEATURES

source

1..1841
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/map="92E8-92E8"
1..1841
/gene="CG4288"
/note="alignment with genomic scaffold AB003730"
/db_xref="FLYBASE:FBgn0038799"
133..1641
/gene="CG4288"
/note="Longest ORF"
/codon_start=1
/db_xref="FLYBASE:FBgn0038799"
/product="GH23975p"
/protein_id="AAL28324.1"
/db_xref="GI:16768210"

CDS

/translation="MTDLESPKKNKHLQVHONNNEITSEPTLWFRNKRORYVVL
AFEPFNYSILRVNLSVAIVAMTENTVYDAGQVSYODFPDMSKOKLILSFEYG
VITLOFGVIGYIGKNTVFGIGISTVILTLTPMAHSHLEMTLFRITIEGPEG
VPRGIIHAWMARMSPLERSRMASTPACNAYGATVYAMCSEFLAKYWMESVYVFG
TTGVITWTIMLVFKAGPELDRECSKECDYLOKRTIGYGSKVRKHPRAITSMPT
ALWASHFSENWGYTLTLPDPSFLRDTLNFEDKGTGILSAVPLAMGILLAVSGYIAD
WLOKGIWTTTOVRNFGAFIAQVFMMLAYLLDPWVSYSLLIAGLAFAMSG
FAVNHLDIADPOHNASVLMGIGNTFATIPGIVSPLLTGVTVTNOTSDMRITFIISAGIY
LVGCVTYMYPCSDLOEMAKTPQKQOEAEKNAQLDTQTAGFVNSGALMKD"

BASE COUNT 392 a 495 c 479 g 475 t

ORIGIN

alignment_scores:

Quality: 992.00 Length: 448
Ratio: 2.892 Gaps: 3
Percent Similarity: 76.562 Percent Identity: 42.411

alignment_block:

us-09-776-865-2 x AY060776 ..

Align seg 1/1 to: AY060776 from: 1 to: 1841

```

80 ArgTyraSnuLeuAlaIleLeuAlaPhePheGlyPhePheIleValTyraI 96
|||||
244 CGGTACATTTGGTGGCTACTGCTTTCTTCGATTCCTTCATGTCTACTC 293
|||||
96 AleuArgValAsnleuSerValAlaIleValAlaIleValAlaIleValAla 113
|||||
294 ACTGGAGTGAATCTTTCGGTGGCATTTGTCGATGACGACAGATCGCA 343
|||||

```


TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
Direct Submission
Submitted (17-FEB-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 167928)

REFERENCE
AUTHORS

Wheeler, D., Weinstock, G., Gibbs, R. and Venter, J.C.
Unpublished
2 (bases 1 to 167928)
Worley, K.C.
Direct Submission
Submitted (01-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 167928)
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alstbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbarta, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, T., Kovar, C., Kratovic, U., Kuresh, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogund, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Ruitz, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabot, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
Direct Submission
Submitted (01-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 167928)
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alstbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbarta, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, T., Kovar, C., Kratovic, U., Kuresh, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogund, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Ruitz, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabot, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
Direct Submission
Submitted (29-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 167928)
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alstbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbarta, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, T., Kovar, C., Kratovic, U., Kuresh, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogund, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Ruitz, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabot, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
Direct Submission
Submitted (29-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 167928)
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alstbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbarta, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Hollway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, T., Kovar, C., Kratovic, U., Kuresh, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogund, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Ruitz, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabot, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
Direct Submission
Submitted (01-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
6 (bases 1 to 167928)
Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alstbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbarta, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C.,


```

150995 TCTGTGGCTTCCGTCCGTATGTCATCGTCTAATCATCATCATCCG 151044
492 AaThVaTgLyGlnTrpGlnThrValPheTyrlleAlaAlaLeuS 508
:::|||||::: ::|||::: |||||::: |||||::: |||||::: ||
151045 GAGACTCTACACAGTGCATCTGCTCTTGGCTGGCGGGCTTGAA 151094
508 nValPheGlyAlaIlePhePheThrPheAlaTyAlaTyGluValGlnA 525
::: ||| ::|||::: ::|||::: |||||::: |||||::: |||
151095 TATAGCGGCTACTTACTTACTTACTTCTTCCGACGCCGAGACCAA 151144
525 snTrpAla 527
:::|||||:::
151145 GCTGGTCG 151152

seq_name: gb_in:AC023711
seq_documentation_block:
LOCUS AC023711 174163 bp DNA linear INV 19-DEC-2001
DEFINITION Drosophila melanogaster 3 BAC RP98-6C4 (Rosewell Park Cancer
Institute Drosophila BAC library) complete sequence.
ACCESSION AC023711
VERSION AC023711.3 GI:17933778
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 174163)
Munry,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,
Gocayne,J.D., Taber,P., Williamson,A., Homs1,F.H.,
Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
Ayele,M.A., Scott,G.S., Worley,K.W., Amaratilake,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,
Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,
Diaperport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
Dwyer,H., Emery-Cohen,A., Ferriter,S., Garg,N.D.S., Houck,J.,
Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,
Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Piltman,G.S., Puri,V.,
Schaefer,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,
Xiang,J., Zaveri,J.S., Zhou,Y., Zorrilla,S., Smith,H.O.,
Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.
Direct Submission
Unpublished
2 (bases 1 to 174163)
Worley,K.C.
Direct Submission
Submitted (17-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 174163)
Worley,K.C., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amratunge,H.C., Are,J.R., Banks,T., Barbarta,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Breva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoque,M., Holloway,C.,
Hollins,B., Homs1,F., Howard,S., Huber,J., Hulik,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kovar,J.,
Kovar,C., Kralovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,C., Liu,J., Liu,W.,

```

```

TITLE
JOURNAL
COMMENT
Submitted (19-DEC-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 19, 2001 this sequence version replaced gi:6997282.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu

```

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STS are identified using ePCR (genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences. Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarly (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

```

FEATURES
source
QUALSTAT-REPORT.
Location/Qualifiers
1..174163
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3"
/clone="RP98-6C4"
BASE COUNT 51926 a 36149 c 35965 g 50523 t
ORIGIN

```

```

alignment_scores:
Quality: 984.50 Length: 486

```

Ratio: 2.921 Gaps: 8
Percent Similarity: 69.342 Percent Identity: 40.947

alignment block:

US-09-776-865-2 x AC023711 ..

Align seg 1/1 to: AC023711 from: 1 to: 174163

```

80  ArgTrpAsnLeuAlaIleLeuAlaPhephegIlyPhepheIleValTYRAl 96
    |||:::  ::::::::::::::::::::|||:::|||||
120238  CGTCAcATTTTCGATTCATCGGCTTCCTGGGATTCGCCCGTGGCTCAcGC 120347
    96  aLeuArgValAsnLeuSerValAlaLeuValAspMetValAsp..... 110
    |||:::|||||:::|||||:::|||||:::|||||:::
120348  GATCGGGGTCAATCTGTGGCGCATTTGTGCCATGTGCATGTGAACCAACGG 120397
    111  ::::::::::::SerAsnThrThrLeuGluAspAsnArgThrSerLysAla 123
    |||:::|||||:::|||||:::|||||:::|||||
120398  CAATTCGGCACAGCAcTCAcTGTGATTTGATTCGAGCAcG..... 120438
    124  CysProGluHisSerAlaProIleLysValHisHisAsn..... 136
    |||:::|||||:::|||||:::|||||:::|||||
120439  TGTCA...CTACCGGCACCA.....CATCAcAATGCTAGCAGATCC 120476
    137  ::::::::::GlnThrGlyLysLysTrpGlnTrpAspAlaGluThrGlnGly 151
    |||:::|||||:::|||||:::|||||:::|||||
120477  CAATCCGAGAAAGAGGAGGAGTTGTGTGGACAGAGCCACGAGGAGAT 120526
    151  rPileLeuGlySerPhePheTrpGlyTrpIleIleThrGlnIleProGly 167
    ::::::::::|||:::|||||:::|||||:::|||||:::|||||
120527  TGGTGCCTGGCAGATTCTTCTATGCTATGTGCTAACCCCAAGTCCCGGC 120576
    168  GlyTrpValAlaSerLysIleGlyLysMetLeuLeuGlyPheGlyIle 184
    |||:::|||||:::|||||:::|||||:::|||||
120577  GGAGCGATGGCGCGAGCTGTATGGTGGAGAAAGATCTACGCGTATGAGAT 120626
    184  eLeuGlyThrAlaValLeuThrIleuPheThrProIleAlaAlaAspLeuG 201
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
120627  GTTGATCAcCGCGGCTTTACGCTTATACCTCAATGTGCTCCCACTGGG 120676
    201  LysAlaGlyProLeuIleValLeuArgAlaLeuGluGlyLeuGlyGlnGly 217
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
120677  ATCTGCCGCTGTGTGCTGTGCTCCGATCTCTGGAGGAAATGGGAGAGGC 120726
    218  ValThrPheProAlaMetHisAlaMetTrpSerSerTrpAlaProPhe 234
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
120727  GTCAcCATTCACAGCTATGCAcCGCATGCTTGCCACCTGAGATACCGCGCT 120776
    234  uGluArgSerLysLeuLeuSerIleSerTrpAlaGlyAlaGlnLeuGly 251
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
120777  GGAAGAGAAcAAGTTCCCGCAATCTGTATGCGGCTCCCAATATCGGAA 120826
    251  hValIleSerLeuProLeuSerGlyIleIleCys.....TyrTyr 264
    |||:::|||||:::|||||:::|||||:::|||||
120827  CAGTCATTTCCATCCCGCTGGCCGAGATGCTGTCTGCTGGAGACTTCTGT 120876
    265  MetAsnTrpThrTrpValPheTrpPhePheGlyThrIleGlyIlePheTr 281
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
120877  GGTGGCTGGCGCGCTGCTTCTCACTCTGCACTGTGGGCACTTCGTG 120926
    281  pPheLeuLeuTrpIleTrpLeuValSerAspThrProGlnLysHisLys 298
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
120927  GTTCATCTGCATCGATGTATTTGGTGTACGAAAGCCTAAGCATCATCCA 120976
    298  rGlieSerHisTrpGlyLysGlyTrpIleLeuSerSerLeuArgAsnGln 314
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
120977  GGATCTCAGATCGAGCGAGAGTATTCGAAAGAGTCTACACAGTTTCAG 121026
    314  ::::::::::::::::::::
121027  AGGCTAATAATCAcGATCTAGCGGAGCCGAGAGAAAGAGAGGACAGCA 121076
    315  ::::::::::::::::::::LeuSerSerGlnLysSerValProTrpValProI 326

```

```

121077  TGAAGTAGTCTGGCGGCGCGCGGAGAAcCATACCTCGGTCAcTGC 121126
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
326  lLeuLysSerLeuProLeuTrpAlaIleValAlaHisPheSerTrp 342
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
121127  TGTCTACATCTCGTAcTCTGTGGGCAcTGTGTGAGCAATCGCGCAC 121176
    343  AsnTrpThrPheTrpThrLeuLeuThrLeuLeuProThrTrpMetLysG 359
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
121177  GGATGGGCGCTTCACAGCGACGATACCGAGAGTCCGCCACTACATGAGCA 121226
    359  uIleLeuArgPheAsnValGlnGluAsnGlyPheLeuSerSerLeuProT 376
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
121227  CATCTCAcATTTTGACATTCACATGCAATGCTGTGCTCAATGCGGTGCG 121276
    376  yTrpLeuGlySerTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 392
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
121277  ATCTAACCTCTCTGTGTGTGGGCAcTGTGCTGCTGCCCGCTGGCGGAT 121326
    393  LeuArgAlaLysTrpAsnPheSerThrLeuCysValArgArgIlePhe 409
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
121327  ATGCTACCCAGACCGCTACATATCCTGCTGCACTCGTATAGTTGTGGAA 121376
    409  rLeuIleGlyMetIleGlyProAlaValPheLeuValAlaAlaGlyPhe 426
    ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
121377  CACGTGCGCTCGGTGGTGGCTCATCACTGCGCTGATGCGCATCATCATG 121426
    426  lGlyCysAspTrpSerLeuAlaValaPheLeuThrIleSerThrThr 442
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
121427  TGGGCTGCGATTGGGTGGGTGACCTTTATGTTGGCGCGCGTGGCTCT 121476
    443  LeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIleAl 459
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
121477  TTGGCGCGCGCGCGCTATGCTGTGGCAACCAcATGATACATACCTACCT 121526
    459  aProSerTrpAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaThr 476
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
121527  TTCACAGTATGACAGGCGACCATGTATGGCATCACCAATTCGCGCGCAAT 121576
    476  lProGlyMetValGlyPro...ValIleAlaLysSerLeuTrpProAsp 491
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
121577  TCTGGGCTTCTCGGCTCCGTATGCTATCGCTCAATCAcCAATCAATCCG 121626
    492  AsnThrValGlyLysGlnThrValPheTrpIleAlaAlaAlaIleAs 508
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
121627  GAGACCTGTGACACAGTGCATGTGCTTCTGCTGGCGCGCGGCGTTGAA 121676
    508  nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyLysAlaGln 525
    |||:::|||||:::|||||:::|||||:::|||||:::|||||
121677  TATAGCGGTAcTCTCATCTGATCTGATCTTCCGACGCGGAGAGACAA 121726
    525  snTrpAla 527
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
121727  GCTGGTGTG 121734
    seq_name: gb_in:AE003491
    seq_documentation block:
    LOCUS AE003491 323461 bp DNA linear INV 04-OCT-2000
    DEFINITION Drosophila melanogaster genomic scaffold 14200001386053 section 8
    ACCESSION AE003491 AE002593
    VERSION AE003491.2 GI:10728219
    KEYWORDS HTG.
    SOURCE fruit fly.
    ORGANISM Drosophila melanogaster
    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
    Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
    Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
    1 (bases 1 to 323461)
    Adams,M.D., Celisner,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
    Amanatides,P.G., Scherier,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
    George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,

```


Sutton,G.G., Wortman,J.R., Vandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazer,E.G., Helt,G., Nelson,C.R., Gabor
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abri,J.F., Agbayani,A., An,H.J.,
Andrews,Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brotlier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablo,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Kocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischmann,W., Fosler,K., Gabrielian,A.E., Garg,N.S.,
Gelbart,M.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wel,M.H., Idegawa,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kernison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Paclob,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kimms,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svitskas,R., Tector,C., Turner,R.,
Venner,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,X., Zhong,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhou,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006
TITLE 2 (bases 1 to 323461)
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
On Oct 9, 2000 this sequence version replaced gi:7292821.
Location/Qualifiers
1. 323461
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="X"
join(<15214, .15403,15749, .16006,16069, .16312,17722, .18091)
/gene="fine"
/product="CT7563"
/db_xref="FLYBASE:FBan0004396"
/db_xref="FLYBASE:FBgn0040222"
/db_xref="FLYBASE:FBgn0040222"
/protein_id="AAFA48215.2"
/gene="fine"
/note="CG4396"
/map="11D1-11D1"
/db_xref="FLYBASE:FBan0004396"
/db_xref="FLYBASE:FBgn0040222"
join(<15214, .15403,15749, .16006,16069, .16312,17722, .18046)
/gene="fine"
/note="fine gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0004396"
/db_xref="FLYBASE:FBgn0040222"
/protein_id="AAFA48215.2"
/db_xref="GI:10728220"
/translation="MTNAMDIVKNGSANGVDSNDSRTNLIVYLPOTMIOEBS
LSSIGLESCKLVADKVSGLSYGYGVYVRAEDAKRAVNTLNGRLINOKYIKSYA
RSSSEIKGANIYSGLPKRNLSQPDLEGFASFGKITIRILCIDNIGLSKVGIRF
DORNAERAIQDLNGRTGTAEPITVFAFPNPPSSAAQIAPLPLAVITPAAATR
RLAAGALPSAGRIIRYSPLAGDLIANSITLPGNMTGSGWCIFFVYNLAPETEEENLWOLF

PGAVQSVKIRIDLOTSKCKGGEFTMTNDEAVVAIOSLNGCYTGLNRYLOVSFTNK
TKT"
complement(join(<23860, .24239,24538, .25184,25286, .25377,
25443, .25533,26089, .26389,26921, .>27095))
/gene="CG4395"
/product="CT4121"
/db_xref="FLYBASE:FBan0004395"
/db_xref="FLYBASE:FBgn0030437"
complement(<23860, .>27095)
/gene="CG4395"
/map="11D1-11D1"
/db_xref="FLYBASE:FBan0004395"
/db_xref="FLYBASE:FBgn0030437"
complement(join(<23860, .24239,24538, .25184,25286, .25377,
25443, .25533,26089, .26389,26921, .27095))
/gene="CG4395"
/note="CG4395 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0004395"
/db_xref="FLYBASE:FBgn0030437"
/protein_id="AAFA48216.1"
/db_xref="GI:7292823"
/translation="MATTSSDSQNDVVASQATQDNURIFLKLHYACVFRONT
VTDTPSSLSGPAITDSDLPENFSPVRYLENAAMNGCVIDMRVDELAKEBELMA
ITVSATMTATNOKENRFLCPNLNEDGYLCMPRPAGTVALSOYCPDFEGNRLAKETC
LENGSWYRHPVNOTWSNVTNVDYDEDFEOLFIDELVYKVALSLALLISIIIFLG
FLSKCTETRIHVIHFAISLACQAVMIITRYLVRVRSFTIENHNPVRCSEKSYENNTA
LOVAQOLKITINTITFOLMCITGLHYVHYFMVYVFNWCBESLHLHLVLYSYRYS
GNLAIACISLSHMAVSHGRCDLQPLTMDVFTSNLSSESSQWPAIRLRQVMS
GSAVOILTPPLSLISQVFNKVDIVRWKVFIVSWFPIPAIYVFLINVLRIYVRLH
POSQAPAPALIRKAAVRAITLIVPLGLHFLIPYRPGDGTQDHFYOMLSVLYSLOG
FVSLPFCFANHDVFAIRITLNLKLLPSLVPRPPAGSSNGWATTPPSLGLV"
join(<43393, .43627,43701, .43859,44240, .45178,45248, .46407,
47115, .>47153)
/gene="CG15721"
/product="CT35951"
/db_xref="FLYBASE:FBan0015721"
/db_xref="FLYBASE:FBgn0030438"
/db_xref="FLYBASE:FBgn0030438"
/evidence=not_experimental
43393, .>47153
/gene="CG15721"
/map="11D1-11D1"
/db_xref="FLYBASE:FBan0015721"
/db_xref="FLYBASE:FBgn0030438"
/evidence=not_experimental
join(<43393, .43627,43701, .43859,44240, .45178,45248, .46407,
47115, .47153)
/gene="CG15721"
/note="CG15721 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0015721"
/db_xref="FLYBASE:FBgn0030438"
/evidence=not_experimental
/protein_id="AAFA48217.1"
/db_xref="GI:7292824"
/translation="MTAFVLSIGFALLAGARWDYGVYYPAPHYEQOOPYDPCP
SGPESINAGARGAOPSTPSPCEQORRALTGIMWETILNKRCTLVPCDNCDOY
NPVCKYKIDTRNFRSCEBLQVYKCRTRGPKRKHQDGCESRTPVQSRVONTNPYA
INPYPNVNGSRGAVSGQPSGQPSGQPSPOPPYSSSVSPRPYVADPIY
OPYQIWSMNLPEYVAGTSKPYNPYETPTTSTTASTTSTSTSTSTSTSTSTSTST
DKITTPPOGDETEASTVTPNPNRINKVNAVIVDPKSPSTTSTSTSTSTSTSTSTST
SVTTPLRKVIITTAPOHATTAPRSKSLISINAVDPEOEAEKTSNEPTTYKSFN
SYNSASYTTTPAPVTEPTPTKYVNDIIRKINVOISAEPTKADPTQSTLVNSTA
APTNETTLOLPFDKVKINAVESETEKIDEETTKANGCEITPVIGSTASKDETIYTK
SLDKVVKINAVESETEKIDEETHTTYVSTAYVPSVSVYQYDYKGSVVPKADVTYQAP
LEKIVKINAVEINERODEQTEITVASESTSKQPTVPSVPTVSTNVQYTPVYG
STEASKDEETTKSLDKVVKINAVEKEDETKQKAPTYSSVTPYSSASSVTPYVYAS
TEAPKNDETTKMALDKLVKINKAOISTEKAVETPTTEPIITTKATPTTYPSYQVYTS
SVYKSSVSSPEDDDEESTRANPLKSIKINAVSDPTQITTPATPTPPTTGSGF
HCVGEPEIDADSVAKSKATSEALPEIVKTESRNIIGTRDQLEW"
complement(join(<47450, .48160,48223, .48387,48516, .48897))
/gene="CG12716"
/product="CT35950"

```
/db_xref="FLYBASE:FBan0012716"
/db_xref="FLYBASE:FBgn0030439"
/evidence=not_experimental
complement(<47450..>48897)
/gene="CG12716"
/map="11D1-11D1"
/db_xref="FLYBASE:FBan0012716"
/db_xref="FLYBASE:FBgn0030439"
/evidence=not_experimental
complement(join(47703..48160,48223..48387,48516..48897))
CDS
/gene="CG12716"
/note="CG12716 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012716"
/db_xref="FLYBASE:FBgn0030439"
/evidence=not_experimental
/protein_id="AA848218.1"
/db_xref="GI:7292825"
/translation="MOTQOKRKYNSRAFTMILLLOVLGILRGVLAOKSPIANWH
MADPHYSDQYAKILGEAGDADDEKHEGLPMLOIYDQPCOEPGVPCATG
TDSFEFNHCRLBAANKMLFOHGTLELEPMEBRCI.PNCOTMCTOVERPVCALEIG
GAIPOEFANCEMRKHECHTKOYLRIIHGPCOTPKSGKKLRKKRPLNLSI
SKRATTPKKVYIMLATPAPRSSSTTPPEFTTTTRITTAARKRPKPTSPSPKRF
QQLMNVANPMVSVQAIIDAYNVNIPDVGHGDRDNLLALDVPARSWSCHRALFRAP
"
mRNA
complement(join(<76685..76911,77099..>77279))
/gene="CG15719"
/product="CT35948"
/db_xref="FLYBASE:FBan0015719"
/db_xref="FLYBASE:FBgn0030440"
/evidence=not_experimental
complement(<76685..>77279)
/gene="CG15719"
/map="11D1-11D3"
/db_xref="FLYBASE:FBan0015719"
/db_xref="FLYBASE:FBgn0030440"
/evidence=not_experimental
complement(join(76685..76911,77099..77279))
CDS
/gene="CG15719"
alignment_scores:
Quality: 984.50 Length: 486
Ratio: 2.921 Gaps: 8
Percent Similarity: 69.342 Percent Identity: 40.947
alignment_block:
US-09-776-865-2 x AE003491 ..
Align seg 1/1 to: AE003491 from: 1 to: 323461
80 ArgTYrAsnLeuAlaIleLeuAlaPhePheGlyPhePheIleValTYrAl 96
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
205408 COTCACTTTTCGATTCATGGGCTTCCTGGGATTCGCCGIGGCTTACGC 205457
96 AleuArYAlaSnLeuSerValAlaLeuValAspMetValAsp 110
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
205458 GATGGGGTCATCTGTCGGTGCATCTGGCCATGGTGAACCAACGG 205507
111 .....SerAsnThrThrLeuGluAspAsnArgThrSerIlySala 123
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
205508 CAATTCCGACAGCAACTCATCGGATTCGATACGACGACG 205548
124 CysProGluHisSerAlaProIleLysValHisHisAsn 136
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
205549 TGTCCA...CTACCGGCACCA.....CATCAAAATGTAAGCATCC 205586
137 .....GlnThrGlyLysLysTYrGlnTrpAspAlaGluThrGlnGly 151
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
205587 CAATCCGAGAAAGAGGCGAGATTGTGTGGGACAGACGACGCGGAT 205636
151 rPleLeuGlySerPhePheTYrGlyTYrIleIleThrGlnIleProGly 167
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
205637 TGTGCTCGGAGATTTCCTATAGGCTATGTGCTAACCAAGTCCCGCG 205686
```

```
168 GlyTYrValAlaSerLysIleGlyLysMetLeuLeuGlyPheGlyI 184
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
205687 GCACGAGATGCCCGACCTGATAGTGGGAAGAAGATCTACGGCTATGGAGT 205736
184 eLeuGlyThrAlaValLeuThrLeuPheThrProIleAlaAspLeuG 201
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
205737 GTTGATCACGGGGGCTTACGCTTATTAATCCATTCATGAGCTGCCACCTGG 205786
201 lYValGlyProLeuIleValLeuArgAlaLeuGluGlyLeuGlyGlu 217
||| :||| :||| :||| :||| :||| :||| :|||
205787 ATCTCCCGCTGTGGTCTGTGGCCATCTGGAGAGGAATGGCCAGGCG 205836
218 ValTrpPheProAlaMetHisAlaMetTrpSerSerTrpAlaProP 234
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
205837 GTACCTTATCCAGCTATGACAGCCCATGCTTGCCCACTGGATACCGCCCT 205886
234 uGluArgSerLysLeuLeuSerLysSerTYrAlaGlyAlaGlnLeu 251
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
205887 GGAGAGGAAGAAGTTCGCCCAATGCTGATAGCGGGCTCCAAATATCGGAA 205936
251 hrValIleSerLeuProLeuSerGlyIleIleGys.....TYrTYr 264
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
205937 CAGTCATTTCCATGCGCGCTGGCCGATGCGTGTGCTGCGTGCACCTCTCG 205986
265 MetAsnTrpTrpTrpValPheTYrPhePheGlyThrIleGlyIleP 281
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
205987 GGTGGCTGCGCGCTGGCTTTTCTACATCTTCGACACTGGTGGCATTTCTGTG 206036
281 pPheLeuLeuTrpIleTrpLeuValSerAspTrpProGlnLysHis 298
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
206037 GTTCATCCCATGAGATGATTTGGTGTACAGACAAGCTAGCATATCCCA 206086
298 rGlyIleSerHisTYrGlnLysGluTYrIleLeuSerSerLeuArgAsn 314
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
206087 GGATCTCAAAATCGGAGCGAGATATATCGAAAGAGTCTACAGGTTAG 206136
314 ..... 314
206137 AGGCTAATAAATCAAGATCTAGCGGAGCGAGAGAAGAGGAGCAAGCA 206186
315 .....LeuSerSerGlnLysSerValProTrpValProI 326
206187 TGAAGTGAAGTGTGGGGCGCGCGGAGGAACCAATACCTCGTATACGC 206236
326 lLeuLysSerLeuProLeuTrpAlaIleValAlaHisPheSerTYr 342
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
206237 TGTCTACATCCGTACCTCTGTGGCCATCTTGTGACCAATGCGGCCAG 206286
343 AsnTrpTrpPheTYrThrLeuLeuThrLeuLeuProThrTYrMet 359
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
206287 GGATGGGCTTCTACACGACGACCTAACGACGAGCTGCCACCTACATGAGCAA 206336
359 uIleLeuArgPheAsnValGlnGluAsnGlyPheLeuSerSerLeu 376
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
206337 CATCTACACTTTGATGATCCAGTCGAATGCTCTCTCAATGCGGTGCCGT 206386
376 rYrLeuGlySerTrpLeuGysMetIleLeuSerGlyGlnAlaAspAsn 392
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
206387 ATCAAACTCTCTGTGCTGTGGCATTTGCTCTCGCCGCTGGCGGATGG 206436
393 LeuArgAlaLysTrpAsnPheSerThrLeuGysValArgArgIleP 409
||| :||| :||| :||| :||| :||| :||| :|||
206437 ATGCTAGCCACAGCTACATATCCGCTGTGAACTCGATTAAGTTGTGAA 206486
409 rLeuIleGlyMetIleGlyProAlaValPheLeuValAlaIlePhe 426
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
206487 CACGCTGGCTCTCGGTCGTCATCATCTGAGGCTGATGGCATCAATCATG 206536
426 lGlyCysAspTYrSerLeuAlaValAlaPheLeuThrIleSerThr 442
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
206537 TGGGCTGGATTTGGGTGTGGGTCACTTATGTGGCGCGGCTGGGCTCT 206586
```

```

443 LeuGIyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIleAl 459
      :::::::::: ::::: ::::: ::::: :::::
206587 TTCGGCGGCCCTCATCTGCGCAACGATGAATCATACCTGCTCAG 206636
      :::::::::: ::::: ::::: ::::: :::::
459 aProSerTyrAlaGlyIleLeuLeuGlyIlePheAsnThrPheAlaThrI 476
      ::::: ::::: ::::: ::::: :::::
206637 TCCACATATGCGAGGACCATGATGATGCATCACCAATTCGCGCAATA 206686
      ::::: ::::: ::::: ::::: :::::
476 IeProGlyMetValGlyPro...ValIleAlaIleSerLeuThrProAsp 491
      ::::: ::::: ::::: ::::: :::::
206687 TCTGTGCTTCCTCGGCTCCGATGATGATCATCATCATCATCTCCG 206736
      ::::: ::::: ::::: ::::: :::::
492 AsnThrValGlyIleUtrpGlnThrValPheTyrIleAlaIleAlaIleAs 508
      ::::: ::::: ::::: ::::: :::::
206737 GAGACTCTACACAGTCGATCTGCTCTCTGCTGCGCGGCTTGAA 206786
      ::::: ::::: ::::: ::::: :::::
508 nValPheGlyAlaIlePhePheThrPheAlaIleAlaIleGlyIleAla 525
      ::::: ::::: ::::: ::::: :::::
206787 TATAGCGGTAACCTATCTACTGATCTTCGCCAGCGCGAGAGCAAA 206836
      ::::: ::::: ::::: ::::: :::::
525 snTPaIa 527
      ::::: :::::
206837 GCTGCTCG 206844

seq_name: gb_to:AF271235

seq_documentation_block:
LOCUS AF271235 3982 bp mRNA linear ROD 14-JUN-2000
DEFINITION Rattus norvegicus differentiation-associated Na-dependent inorganic
ACCESSION AF271235
VERSION AF271235.1 GI:8515880
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 3982)
AUTHORS Mashima,H. and Kojima,I.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3982)
AUTHORS Mashima,H. and Kojima,I.
TITLE Direct Submission
JOURNAL Submitted (24-MAY-2000) Cell biology, Gunma University, Institute
for Molecular and cellular Regulation, 3-39-15, Showa-machi,
Maebashi, Gunma 371-8512, Japan
FEATURES
Source
1..3982
Location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
1..3982
/gene="DNPI"
463..2211
/gene="DNPI"
/codon_start=1
/product="differentiation-associated Na-dependent inorganic
phosphate cotransporter"
/protein_id="AA076223.1"
/db_xref="GI:8515881"
/translation="MESVKORILAPRKEGIKNFAGKSLGQIVRLERKKQDNRTIELT
EDGKPLEVPEKKAPLDCCTCGLPRLRYIIAISMISGICISGICINGLVAIVDWNNS
TIHKGAVIKERAKFNDPEVGMIHGSEFGYIIIGDYGIVASRLAANVFPAAIL
LSTLNLMLIPSAVAVHYGCYFVRIILOGIVGTVYPAHGHMSKAWAPLESRLATTS
PCGSYACAVIAMPAGIINOTGSSVYVYSGCMYMWMLIYSVPKSPKPTID
ERRRIVESTIESANLIGAMERKFTPMKFTTSPRYAIYANCRSNTFTLLISGP
AYFEVGFETSKVGLSAVPHLWTLIVPIGGIDFELRKSQILSTTVYKILNCGG
FGVETALLVVGYSHTRGVALISFLVAVGSEFASISGFNNVHLDIAPRYASILNGISN
GVTALSGMVCPIIVGAMTKNSREMOYVFILALVHGCVTFYALPASGSEKQWAPD
EETSEKCGFIHEDELDEMGDITQNTINVGTTSYCATSQENSGMWNGMKKEEFG
ESADAVSYKORDYS"
3953..3958
polyA_signal

```

```

BASE COUNT 1144 a 823 c 884 g 1131 t
ORIGIN
alignment_scores:
Quality: 973.00 Length: 502
Ratio: 2.695 Gaps: 11
Percent Similarity: 71.912 Percent Identity: 41.235
alignment_block:
US-09-776-865-2 x AF271235 ..
Align seg 1/1 to: AF271235 from: 1 to: 3982
45 ProValArgAspLeuAlaArgAsnAspGlyIleGluSer..... 57
   ||| ||||| ::::: ||||| |||||
531 CCTCGACAGATCTA...CAGGCTCTGGAGAGAAAGCAGATACCCAG 577
   ::::::::::ThrAspArgThrProLeuProGlyAlaProArgAla 70
58 ..... 71 GluAlaAlaProValCys...CysSerAla.....ArgTyr 81
   ::::: ||||| ::::: |||||
625 AAGAAAGCTCCGCTATGCGACTGCTGCTGCGCTGCGCGCGCTA 674
   ||| ||||| |||||
81 rsnLeuAlaIleLeuAlaPhePheGlyPhePheIleValTyrAlaLeu 98
   ||| ||||| ::::: |||||
675 CATCATAGCCATCATGAGCGGCTCGCTTCGATCTCCCTTGATCC 724
   ::::: ||||| ::::: |||||
98 rValAsnLeuSerValAlaLeuValAspMetValAspSerAsnThrThr 114
   ||| ||||| ::::: |||||
725 GCTGTAACTCGGCTGTGGCATTTGTGACATGTCACAACACACACATC 774
   ::::: ||||| ::::: |||||
115 LeuGluAspAsnArgThrSerTyrAlaCysProGlnHisSerAlaProI 131
   ::::: ||||| ::::: |||||
775 .....CACCGGGGTGGCAAAAGTTATCAAGAGAAAGCC..... 807
131 eLysValHisHisAsnGlnThrGlyLysTyrGlnThrPaspAlaGluT 148
   ||| ||||| ::::: |||||
808 .....AAGTTAACTGGGACCCGAG 829
148 hrcInclYrPrIleLeuGlySerPhePheTyrGlyIleIlePheGln 164
   ||| ||| ||||| ||||| ::::: |||||
830 CTGTGGGATGATTCACGGGCTCTCTCTGCGCATATCATCACAGCAG 879
165 IeProGlyIleTyrValAlaSerTyrIleGlyLysMetLeuLeuG 181
   ||||| ||||| ::::: |||||
880 ATTCGGGGGATACATCGCATCGACGCTGCTAACCGGGCTCTTGG 929
181 yPheGlyIleLeuGlyThrAlaValLeuThrPheThrProIleAla 198
   ||| ::::: ||||| ::::: |||||
930 GGTGTCATACCTCTTACCTCTACCTCAATATGCTGATCCCATCTGCG 979
198 IaAspLeuGlyValGlyProLeuIleValLeuArgAlaLeuGluGlyLeu 214
   || ::::: ||| ::::: |||||
980 CCAGATGCATATGATGATGCGTCATCTTGTATGATATGCAAGACAT 1029
215 GlyGluGlyValThrPheProAlaMetHisAlaMetCtirPseSerTrp 221
   ||||| ||||| ::::: |||||
1030 GTGAGAGGGGTACCTACCAAGCTTCACGGGATVTGAGACAAAGTGCC 1079
231 aProProLeuGluArgSerTyrLeuSerIleSerTyrAlaGlyVal 248
   ||||| ||||| ::::: |||||
1080 CCTCTTTGGAGAGAGTAGTGGCTACCACTCTCTCTGCTGCTTCC 1129
248 IeLeuGlyThrValIleSerLeuProLeuSerGlyIleIleCysTyrTyr 264
   ||| ::::: ||||| ::::: |||||
1130 ATGCTGACAGCATTCGCAATGCCCTTACGTGTCTGCGTACAGTAC 1179
265 MetAsnTrpThrTyrValPheThrPhePheGlyThrIleGlyIlePheThr 281
   ::::: ||||| ::::: |||||

```



```

115 LeuGlnAspAsnAlaThrSerLysAlaCysProGlnHisSerAlaProIl 131
    ::::: ::::: :::::
951 .....CACCGGAGGAAAGTATCAGAGAAAGCC..... 983
131 eLysValHisHisAsnGlnThrLysLysTyrGlnTrpAspAlaLut 148
    ::::: ::::: :::::
984 .....AAATTTAACTGGGAGCCCGAGA 1005
148 hrgInglYrPleLeuGlySerPhePheTyrGlyTyrIleIleThgln 164
    ::::: ::::: :::::
1006 CCGGGGAGATCCAGGATCGTTCTTGGGCTATATCATCACCAG 1055
165 IleProGlyGlyTyrValAlaSerLysIleGlyLysMetLeuLeuG 181
    ::::: ::::: :::::
1056 ATTCACAGAGATATATCGATCGCGCTGCTCAACCGGCTTTGG 1105
181 yPheGlyIleLeuGlyThrAlaValLeuThrPheThrProIleAla 198
    ::::: ::::: :::::
1106 GGCTGCATCTGCTCACCTTACCTCAATATGCTGATCCCATCTCAG 1155
198 IaAspLeuGlyValGlyProLeuIleValLeuAlaLeuGlyLeu 214
    ::::: ::::: :::::
1156 CCAGAGTCATATGATGTGTCTTGTATGATATTCAGAGACTT 1205
215 GlyGlnGlyValThrPheProAlaMetHisAlaMetTrpSerTyrAl 231
    ::::: ::::: :::::
1206 GTGAGGGGTGCACCTACCCAGCTGTATGAGATGAGCAAGTGGGC 1255
231 arProLeuGlnArgSerLysLeuLeuSerLieserTyrAlaGlyAla 248
    ::::: ::::: :::::
1256 CCCGCCCTTGAGAGAGTAGTGGCTACAAACCTCTTGTGGTTCCT 1305
248 lneLysGlyThrValIleSerLeuProLeuSerGlyIleIleCysTyr 264
    ::::: ::::: :::::
1306 ATGTGGAGACGATATTCATGCTTACCTGATCTTGTGCATAC 1355
265 MetAsnTrpThrTyrValPheTyrPhePheGlyThrIleGlyIlePhe 281
    ::::: ::::: :::::
1356 ACTGATGTCGTCAGTATTTATGTATGATGAGACTTGGCAGGCTG 1405
281 pPheLeuLeuThrIleTrpLeuValSerAspThrProGlnHisLys 298
    ::::: ::::: :::::
1406 GTACATCTTCTGCTGCTGTGTCTTATGAGAGCCCGCAAAAGCATCTA 1455
298 rglIleSerHisTyrGlnLysGlnTyrIleLeuSerLeu..... 311
    ::::: ::::: :::::
1456 CCATTACAGATGAGAAAGTACGATACATAGAGAGACATTCGAGAGAC 1505
312 ArgAsnGlnLeuSerSerGlnLys.....SerValProTrpValProIl 326
    ::::: ::::: :::::
1506 GCAAAATCTGCTAGGTGCAATGAGAAATTTAAAGCCCATGAGAGAGTT 1555
326 eleuLysSerLeuProLeuTrpAlaIleValAlaHisPheSerTyr 343
    ::::: ::::: :::::
1556 TTTTCATCTCATGCGCTGACGAGATATGTTGCCAATCTTCGACAGA 1605
343 snrTrpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLysGln 359
    ::::: ::::: :::::
1606 GCTGACATTTTATTTACTGCTCATGTCAGCCAGCCTTATTTGAGAG 1655
360 lLeuLeuArgPheAsnValGlnLysGlnGlyPheLeuSerSerLeuPro 376
    ::::: ::::: :::::
1656 GTTTTTCATTTGAAATCAGCAAGTTGGCATGTTGCTGCAGACCTCA 1705
376 rLeuGlySerTrpLeuLysMetIleLeuSerGlnAlaAlaAspAsnL 393
    ::::: ::::: :::::
1706 CTTTTCATACATCATTTGCTTATCGGGGGGCAATTCGACATTTTCC 1755
393 euaArgAlaLysTrpAsnPheSerThrLeuLysValArgArgIlePheSer 409
    ::::: ::::: :::::
1756 TAAAGACAGCAAAATCTCTCAACAACACTACAGTACAGAAAGATCATGAT 1805
410 LeuIleGlyMetIleGlyProAlaValPheLeuValAlaAlaGlyPheIl 426

```

```

1806 TGTCGGGGTTTGGCATGAGAACCCGCTCTTCTGCTTGGCTAC.. 1853
426 eGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThrTrl 443
    ::::: ::::: :::::
1854 ..TCATACTACAGAGGGGTGGCCATCTCTTGTGCTTGCATACATGAT 1902
443 euGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIleAla 459
    ::::: ::::: :::::
1903 TCAGTGAATTTGCTATCTCTGTTCAATGTTATCTTACCTTGATATGCT 1952
460 ProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaThrIl 476
    ::::: ::::: :::::
1953 CCAAGATATCCACGATATCTTAATGGCATTTCAATGCGCTTGGCAGCT 2002
476 eProGlyMetValGlyProValIleAlaLysSerLeuThrProAspAsn 493
    ::::: ::::: :::::
2003 GTGCGGATGTGTTGGCCCTATCATTTGTTGGTGCATATACAAAGATAGT 2052
493 hrValGlyGlyTrpGlnThrValPheTyrIleAlaIleAlaIleAsnVal 509
    ::::: ::::: :::::
2053 CCCGTGAAGATGCGACTATGCTTCTCCTCATGCTGCACGTCCTCAT 2102
510 PheGlyAlaIlePhePheThrLeuPheAlaLysGlyLysValGlnAsnTr 526
    ::::: ::::: :::::
2103 GGTGAGTCAATATTTATTCACATATTTGCTTGCATGAGAGAAACACTTG 2152
526 pAla 527
    :::::
2153 GGCA 2156
seq_name: gb_pr:AB032435
seq_documentation_block:
LOCUS AB032435 3946 bp mRNA linear PRI 18-DEC-2001
DEFINITION Homo sapiens DNPI mRNA for differentiation-associated Na-dependent
inorganic phosphate cotransporter, complete cds.
ACCESSION AB032435
VERSION AB032435.1 GI:7328922
KEYWORDS differentiation-associated Na-dependent inorganic phosphate
cotransporter.
SOURCE Homo sapiens brain thalamus cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (sites)
Aihara,Y., Mashima,H., Onda,H., Hisano,S., Kasuya,H., Hori,T.,
Yamada,S., Tomura,H., Yamada,Y., Inoue,I., Kojima,I. and Takeda,J.
Molecular cloning of a novel brain-type Na(+)-dependent inorganic
phosphate cotransporter
J. Neurochem. 74 (6), 2622-2625 (2000)
20281869
JOURNAL MEDLINE
2 (bases 1 to 3946)
Takeda,J. and Onda,H.
AUTHORS Direct Submission
TITLE Submitted (15-SEP-1999) Jun Takeda, Gunma University, Institute for
Molecular and Cellular Regulation, Laboratory of Molecular
Genetics, Department of Cell Biology, 3-39-15, Showa-machi,
Gunma 371-8512, Japan (E-mail:j.takeda@akagi.sb.gunma-u.ac.jp,
Tel:81-27-220-8830, Fax:81-27-220-8889)
LOCATION/Qualifiers
FEATURES
source
1..3946
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="brain thalamus"
/note="vector:pGEM3zf(+); site1,2:ECORI"
414..2162
/gene="DNPI"
414..2162
/gene="DNPI"
/codon_start=1
/product="differentiation-associated Na-dependent
inorganic phosphate cotransporter"

```


523 ValGlnAsnTrpAla 527
||| |||||
1917 AAACAACCCCTGGCA 1931

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 08:22:15 ; Search time 53.59 seconds

(without alignments)
1730.272 Million cell updates/sec

Title: US-09-776-865-2

Perfect score: 2836

Sequence: 1 MAAGAMTPRPVQPARPGF.....LFAKGEVQNALNDHGRH 536

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP-archaea:*

2: SP-bacteria:*

3: SP-fungi:*

4: SP-human:*

5: SP-invertebrate:*

6: SP-mammal:*

7: SP-mhc:*

8: SP-organellar:*

9: SP-phage:*

10: SP-plant:*

11: SP-rodent:*

12: SP-virus:*

13: SP-vertebrate:*

14: SP-unclassified:*

15: SP-virus:*

16: SP-bacteriap:*

17: SP-archaeap:*

SUMMARIES

| Result No. | Score | Query | Length | DB | ID | Description |
|------------|-------|-------|--------|----|--------|--------------------|
| 1 | 2836 | 100.0 | 536 | 4 | Q9NRA2 | Q9NRA2 homo sapien |
| 2 | 2621 | 92.4 | 495 | 4 | Q9UGH0 | Q9UGH0 homo sapien |
| 3 | 2329 | 82.1 | 495 | 6 | Q9WMD1 | Q9WMD1 ovis aries |
| 4 | 996.5 | 35.1 | 559 | 5 | Q9VYG7 | Q9VYG7 drosophila |
| 5 | 992 | 35.0 | 502 | 5 | Q9VDM0 | Q9VDM0 drosophila |
| 6 | 970 | 34.2 | 582 | 11 | Q9J112 | Q9J112 rattus norv |
| 7 | 970 | 34.2 | 582 | 11 | Q920B7 | Q920B7 mus musculu |
| 8 | 965.5 | 34.0 | 582 | 4 | Q9P208 | Q9P208 homo sapien |
| 9 | 946.5 | 33.4 | 586 | 5 | Q23514 | Q23514 caenorhabd |
| 10 | 941 | 33.2 | 560 | 4 | Q9P2U7 | Q9P2U7 homo sapien |
| 11 | 936 | 33.0 | 560 | 11 | Q62634 | Q62634 rattus norv |
| 12 | 928.5 | 32.7 | 529 | 5 | Q9Y7S5 | Q9Y7S5 drosophila |
| 13 | 917 | 32.3 | 483 | 5 | Q61369 | Q61369 drosophila |
| 14 | 895 | 31.6 | 466 | 4 | Q96LH1 | Q96LH1 homo sapien |
| 15 | 876 | 30.9 | 497 | 4 | Q9Y2C5 | Q9Y2C5 homo sapien |
| 16 | 856 | 30.2 | 479 | 5 | Q9W4G7 | Q9W4G7 drosophila |

| | | | | | | |
|----|-------|------|-----|----|--------|--------------------|
| 17 | 852.5 | 30.1 | 496 | 5 | Q9VKC9 | Q9VKC9 drosophila |
| 18 | 838.5 | 29.6 | 481 | 5 | Q9VJW8 | Q9VJW8 drosophila |
| 19 | 822.5 | 29.0 | 453 | 5 | Q9NKP8 | Q9NKP8 drosophila |
| 20 | 821.5 | 29.0 | 512 | 5 | Q9VPX2 | Q9VPX2 drosophila |
| 21 | 814.5 | 28.7 | 465 | 11 | Q91WV5 | Q91WV5 mus musculu |
| 22 | 811 | 28.6 | 493 | 5 | Q9VRA4 | Q9VRA4 drosophila |
| 23 | 808 | 28.5 | 563 | 5 | Q9TZN7 | Q9TZN7 caenorhabd |
| 24 | 798.5 | 28.2 | 467 | 4 | Q9H531 | Q9H531 homo sapien |
| 25 | 798 | 28.1 | 465 | 5 | Q9V905 | Q9V905 drosophila |
| 26 | 797.5 | 28.1 | 573 | 5 | Q09932 | Q09932 caenorhabd |
| 27 | 782.5 | 27.6 | 497 | 5 | Q9V8M8 | Q9V8M8 drosophila |
| 28 | 762 | 26.9 | 475 | 5 | Q9V8N0 | Q9V8N0 drosophila |
| 29 | 762 | 26.9 | 491 | 5 | Q9V8N1 | Q9V8N1 drosophila |
| 30 | 762 | 26.9 | 560 | 5 | Q9VOC0 | Q9VOC0 drosophila |
| 31 | 760 | 26.8 | 529 | 10 | Q9SD14 | Q9SD14 oryza sativ |
| 32 | 756.5 | 26.7 | 462 | 5 | Q9V8M9 | Q9V8M9 drosophila |
| 33 | 747 | 26.3 | 512 | 10 | Q82390 | Q82390 arabidopsis |
| 34 | 729 | 25.7 | 449 | 5 | Q9V334 | Q9V334 drosophila |
| 35 | 699.5 | 24.7 | 390 | 11 | Q921B2 | Q921B2 mus musculu |
| 36 | 696.5 | 24.6 | 524 | 5 | Q94886 | Q94886 drosophila |
| 37 | 696.5 | 24.6 | 524 | 5 | Q9V763 | Q9V763 drosophila |
| 38 | 643 | 22.7 | 495 | 5 | Q9W120 | Q9W120 drosophila |
| 39 | 634 | 22.4 | 527 | 5 | Q16923 | Q16923 caenorhabd |
| 40 | 626.5 | 22.1 | 420 | 4 | Q9H533 | Q9H533 homo sapien |
| 41 | 583.5 | 20.6 | 444 | 5 | Q9W121 | Q9W121 drosophila |
| 42 | 582 | 20.5 | 468 | 5 | Q9V9J0 | Q9V9J0 drosophila |
| 43 | 569 | 20.1 | 792 | 5 | Q9W123 | Q9W123 drosophila |
| 44 | 566.5 | 20.0 | 432 | 10 | Q9FKV1 | Q9FKV1 arabidopsis |
| 45 | 562 | 19.8 | 413 | 10 | Q23065 | Q23065 arabidopsis |

ALIGNMENTS

RESULT 1
ID Q9NRA2 PRELIMINARY; PRT; 536 AA.
AC Q9NRA2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE MEMBRANE GLYCOPROTEIN HP59.
GN HP59.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Fu C., Bardhan S., Cetateanu N.D., Lloyd S.R., Yan H., Carter C.E.,
RA Shi E., Venkov C., Yakes M.F., Page D.L., Hellerqvist C.G.; that
RT Identification of a novel membrane protein from mammalian cells that
RT interacts with the anti-pathoangiogenic compound CM101."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF244577; AAF97769.1; -
SQ SEQUENCE 536 AA; 58721 MW; 3DDADFC84007724 CRC64;

Query Match 100.0%; Score 2836; DB 4; Length 536;
Best Local Similarity 100.0%; Pred. No. 1e-217;
Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAAGAMTPRPVQPARPGFGLSGRSLICQVASTPAHGVMSRPVRLARDNDESDTR 60
DB 1 MAAGAMTPRPVQPARPGFGLSGRSLICQVASTPAHGVMSRPVRLARDNDESDTR 60
QY 61 TPLLPAPRAEAPVCCSRYNLAIAIFGFFIVYALRVNLGVALVDWDSMTTLEDNPT 120
DB 61 TPLLPAPRAEAPVCCSRYNLAIAIAIFGFFIVYALRVNLGVALVDWDSMTTLEDNPT 120
QY 121 SKACPEHSAPIKVHNHQTQKTYQMDAETQGWILGFFYYITQJPGYVASKIGGKML 180
DB 121 SKACPEHSAPIKVHNHQTQKTYQMDAETQGWILGFFYYITQJPGYVASKIGGKML 180

```

QY 181 GFGILGTAVALTFTPIADLGVGPLIVLRALLEGEGVTFPAMHAMSSNAPLERSKL 240
    |||
DB 181 GFGILGTAVALTFTPIADLGVGPLIVLRALLEGEGVTFPAMHAMSSNAPLERSKL 240
QY 241 SISVAGAQGLTVISLPLSGIICYYMMNTYVFFPGTIGIFWELLIMLVSDTPQKHRS 300
    |||
DB 241 SISVAGAQGLTVISLPLSGIICYYMMNTYVFFPGTIGIFWELLIMLVSDTPQKHRS 300
QY 301 HYEKEYLLSLRNQSSOKSVPMWPIKSLPLMAIVVAHESYMMTEVTLTLPTMYKEI 360
    |||
DB 301 HYEKEYLLSLRNQSSOKSVPMWPIKSLPLMAIVVAHESYMMTEVTLTLPTMYKEI 360
QY 361 LRFVVOENGFLSLPLYLGSWMLCMLISGQADNLRANKNFSTLCRRIFSLIGMIGPAVFL 420
    |||
DB 361 LRFVVOENGFLSLPLYLGSWMLCMLISGQADNLRANKNFSTLCRRIFSLIGMIGPAVFL 420
QY 421 VAAFGICDYSLAVAFITITTLGFCSSGFSINHLDIAPSVAGILGINTFATIPGMV 480
    |||
DB 421 VAAFGICDYSLAVAFITITTLGFCSSGFSINHLDIAPSVAGILGINTFATIPGMV 480
QY 481 GPVIAKSLTPEDNTEGEMQTVFYIAAIVNFGAIFFTLFARGEVOMALNDHGHHRH 536
    |||
DB 481 GPVIAKSLTPEDNTEGEMQTVFYIAAIVNFGAIFFTLFARGEVOMALNDHGHHRH 536

```

```

RESULT 2
ID 09UGHO PRELIMINARY; PRT; 495 AA.
AC 09UGHO;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE STALIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20047778; PubMed=10581036;
RA Verheijen F.W., Verbeek E., Aulá N., Beereus C.E.M.T., Havelaar M.C.,
RA Joosse M., Peltonen L., Aulá P., Galjaard H., Van der Spek P.J.V.D.,
RA Mancini G.M.S.;
RT "A new gene, encoding an anion transporter, is mutated in sialic acid
RT storage diseases.";
RL Nat. Genet. 23:462-465(1999).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL; AJ387747; CAB62540.1; -.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr.1.
KW Transmembrane.
SQ SEQUENCE 495 AA; 54639 MW; 5C6C15AB3E93A19E CRC64;

```

Query Match 92.4%; Score 2621; DB 4; Length 495;
 Best Local Similarity 100.0%; Pred. No. 1.2e-200;
 Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 42 MRSFVRLDARNDEESTDRTPLLPGARBARADAVCCSARYNLAILAFGFFIYALRVNL 101
    |||
DB 1 MRSFVRLDARNDEESTDRTPLLPGARBARADAVCCSARYNLAILAFGFFIYALRVNL 60
QY 102 SVLVLMVDSNTLTLENDRTSKACREHSAPRKVHNHNOTGKKYQWDAETQGLISFFGYI 161
    |||
DB 61 SVLVLMVDSNTLTLENDRTSKACREHSAPRKVHNHNOTGKKYQWDAETQGLISFFGYI 120
QY 162 ITQIPGGYVASKIGKMLGFGILGTAVALTFTPIADLGVGPLIVLRALLEGEGVTFP 221
    |||
DB 121 ITQIPGGYVASKIGKMLGFGILGTAVALTFTPIADLGVGPLIVLRALLEGEGVTFP 180
QY 222 AMHAMSSNAPLERSKLSISVAGAQGLTVISLPLSGIICYYMMNTYVFFPGTIGIFW 281

```

```

DB 181 AMHAMSSNAPLERSKLSISVAGAQGLTVISLPLSGIICYYMMNTYVFFPGTIGIFW 240
    |||
QY 282 FILMIMLVSDTPQKHRSIHYEKEYLLSLRNQSSOKSVPMWPIKSLPLMAIVVAHFS 341
    |||
DB 241 FILMIMLVSDTPQKHRSIHYEKEYLLSLRNQSSOKSVPMWPIKSLPLMAIVVAHFS 300
QY 342 YNNTFYLLTLPTMYKEILRFVVOENGFLSLPLYLGSWMLCMLISGQADNLRANKNFST 401
    |||
DB 301 YNNTFYLLTLPTMYKEILRFVVOENGFLSLPLYLGSWMLCMLISGQADNLRANKNFST 360
QY 402 LCVRRIFSLIGMIGPAVFLVAAFGICDYSLAVAFITITTLGFCSSGFSINHLDIAP 461
    |||
DB 361 LCVRRIFSLIGMIGPAVFLVAAFGICDYSLAVAFITITTLGFCSSGFSINHLDIAP 420
QY 462 YAGILGINTFATIPGMVGPVIAKSLTPEDNTEGEMQTVFYIAAIVNFGAIFFTLFARG 521
    |||
DB 421 YAGILGINTFATIPGMVGPVIAKSLTPEDNTEGEMQTVFYIAAIVNFGAIFFTLFARG 480
QY 522 EVOQWALNDHGHHRH 536
    |||
DB 481 EVOQWALNDHGHHRH 495

```

```

RESULT 3
ID 09MZD1 PRELIMINARY; PRT; 495 AA.
AC 09MZD1;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE MEMBRANE GLYCOPROTEIN SP55.
GN SP55.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Fu C., Bardhan S., Cetateanu N.D., Lloyd S.R., Yan H., Carter C.E.,
RA Shi E., Venkov C., Yakes M.F., Page D.L., H.C.G.;
RT "Identification of a novel membrane protein from mammalian cells that
RT interact with the anti-pathogenic compound CM101.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF244578; AAF97770.1; -.
SQ SEQUENCE 495 AA; 54536 MW; 649D7CA459B28272 CRC64;

```

Query Match 82.1%; Score 2329; DB 6; Length 495;
 Best Local Similarity 86.3%; Pred. No. 2.3e-177;
 Matches 427; Conservative 39; Mismatches 29; Indels 0; Gaps 0;

```

QY 42 MRSFVRLDARNDEESTDRTPLLPGARBARADAVCCSARYNLAILAFGFFIYALRVNL 101
    |||
DB 1 MRSFVRLDARNDEESTDRTPLLPGARBARADAVCCSARYNLAILAFGFFIYALRVNL 60
QY 102 SVLVLMVDSNTLTLENDRTSKACREHSAPRKVHNHNOTGKKYQWDAETQGLISFFGYI 161
    |||
DB 61 SVLVLMVDSNTLTLENDRTSKACREHSAPRKVHNHNOTGKKYQWDAETQGLISFFGYI 120
QY 162 ITQIPGGYVASKIGKMLGFGILGTAVALTFTPIADLGVGPLIVLRALLEGEGVTFP 221
    |||
DB 121 ITQIPGGYVASKIGKMLGFGILGTAVALTFTPIADLGVGPLIVLRALLEGEGVTFP 180
QY 222 AMHAMSSNAPLERSKLSISVAGAQGLTVISLPLSGIICYYMMNTYVFFPGTIGIFW 281
    |||
DB 181 AMHAMSSNAPLERSKLSISVAGAQGLTVISLPLSGIICYYMMNTYVFFPGTIGIFW 240
QY 282 FILMIMLVSDTPQKHRSIHYEKEYLLSLRNQSSOKSVPMWPIKSLPLMAIVVAHFS 341
    |||
DB 241 FILMIMLVSDTPQKHRSIHYEKEYLLSLRNQSSOKSVPMWPIKSLPLMAIVVAHFS 300

```

| | | | |
|--------|---|--|---------|
| Qy | 342 | YNNFFYLLTLPLPRYKELLRFVNOEKGFLSSPLPYGSLMCLTSLSQADNIRAKNNST | 401 |
| Db | 301 | YNNFFYLLTLPLPRYKELLRFVNOEKGFLSSPLPYGSLMCLTSLSQADNIRAKNNST | 360 |
| Qy | 402 | LCVRIEISLGMIGPAFLVAAGFIGCDYSLAAVFLTISTTGGFCSSGFSINHLDIAPS | 461 |
| Db | 361 | LMVRRVSLGMIGPAFLVAAGFIGCDYSLAAVFLTISTTGGFCSSGFSINHLDIAPS | 420 |
| Qy | 462 | YAGLLIGITNTFATIPGMVGPVIAKSLTPDNTVGENQVFFYIAAALNFGAIFTLFAKG | 521 |
| Db | 421 | YAGLLIGITNTFATIPGMIGPIIARSLTPENTIGEMQVFCIAAALNFGAIFTLFAKG | 480 |
| Qy | 522 | EVQNNALNDHGHHRH | 536 |
| Db | 481 | EVQNNALSDHGHHRN | 495 |
| RESULT | 4 | | |
| 09YVG7 | | | |
| ID | 09YVG7 | PRELIMINARY; | 559 AA. |
| AC | 09YVG7 | | |
| DT | 01-MAY-2000 (TrEMBLrel. 13, Created) | | |
| DT | 01-MAY-2000 (TrEMBLrel. 13, Last sequence update) | | |
| DT | 01-JUN-2001 (TrEMBLrel. 17, Last annotation update) | | |
| DE | CG4330 PROTEIN. | | |
| CN | CG4330. | | |
| OS | Drosophila melanogaster (Fruit fly). | | |
| OC | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; | | |
| OC | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | | |
| OC | Ephydroidea; Drosophilidae; Drosophila. | | |
| CC | NCBI_TaxID=7227; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=BERKELEY; | | |
| RX | MEDLINE=20196006; PubMed=10731132; | | |
| RA | Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., | | |
| RA | Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., | | |
| RA | George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., | | |
| RA | Stuton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., | | |
| RA | Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D., | | |
| RA | Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., | | |
| RA | Abbel J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., | | |
| RA | Baltes R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., | | |
| RA | Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., | | |
| RA | Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P., | | |
| RA | Burtis K.C., Busam D.A., Butler H., Cadieux E., Center A., Chandra I., | | |
| RA | Cherry J.M., Cawley S., Dahlke S., Davenport L.B., Davies P., | | |
| RA | de Pallos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M., | | |
| RA | Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., | | |
| RA | Durbin K.J., Evangelista C.C., Ferraz S., Ferrieres S., Fleischmann W., | | |
| RA | Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., | | |
| RA | Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., | | |
| RA | Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., | | |
| RA | Hoskins D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C., | | |
| RA | Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., | | |
| RA | Kimmel B.E., Kodira C.D., Kraft C., Krautitz S., Kulp D., Lai Z., | | |
| RA | Liao X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., | | |
| RA | Lin X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., | | |
| RA | Montuslov G., Mlshina N.V., Moadary C., Morris J., Moshrefi A., | | |
| RA | Mourou G.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., | | |
| RA | Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M., | | |
| RA | Palazzo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G., | | |
| RA | Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., | | |
| RA | Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., | | |
| RA | Spieler E., Spradling A.C., Stapleton M., Strong R., Sun E., | | |
| RA | Svaystas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., | | |
| RA | Wang Z.-T., Wassarman D.A., Weinstock G.M., Weissbach J., | | |
| RA | Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., | | |
| RA | Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., | | |
| RA | Zheng X.H., Zhong F., Zhong W., Zhou X., Zhu S., Zhu X., Zhang H.O., | | |
| RA | Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; | | |
| RT | "the genome sequence of Drosophila melanogaster."; | | |

```

RL   Science 287:2185-2195(2000).
CC   -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL: AE003491; AAF48230.1; -.
DR   Flybase: FBgn0030452; CG4330.
DR   InterPro: IPR003662; sub_transporter.
KW   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
SQ   SEQUENCE 559 AA; 61744 MW; 8C48402881046C70 CRC64;

Query Match          35.1%; Score 996.5; DB 5; Length 559;
Best Local Similarity 39.8%; Pred. No. 3.7e-71;
Matches 208; Conservative 91; Mismatches 171; Indels 53; Gaps 10;

OY   51 RNDGESTDR-----PRLPGAPRAAPRC-CSARYNALIALEFFGFYALRVNLS 102
DB   31 RSDDDADDERERFCGGERPLRSSGAALENHCCKTKHIFGFMGLFQAVYAARVLS 90
OY   103 VALYMDV-----SNTTLEDNRITSKACREHSAPIKVHN-----QTKKKYQWDAETOG 150
DB   91 VAIVAMNOTALPHNSNSVIDPDT---CP-LPAP---HHNSDPRPQKEGEFVMDAETOG 143
OY   151 WILGSEFFGIYITQIPGVASKIGCKMKLGFGLIGTVLFLFPPIAADLQVGLVYRA 210
DB   144 IYLGSFFGYVLYVPGGRMAELYGKKIYGIVGLITAVFTLPLAHHMDPLPLVRI 203
OY   211 LEGEGEYTPRPMHMMSSMAPLERSKLSTISYVAGDAGVYSLPGIIC---YUNMW 267
DB   204 LEGMGEGYTPRPMHMLHWPDERLNFAIVAGSINGIVISMPPLAGMLCSLDFLGG 263
OY   268 TVYVFFFTIGIFWPLIMWLVSDDPQKHRIKISHKEKYILSSLRNQ----- 314
DB   264 PSAEYIFGLIGIMFIAMWLYVQKPSDHPRISESEERYIERSLDVQVRLINDLAEE 323
OY   315 -----LSQKSPVNPVPIKSLPLMAIVAHFSYMWTEYTLTLPLTYMKETLRFNV 365
DB   324 EGQDEVSLRAPPEEPIPMSSILTSVPLMAILLTCCGCGMAFYTQLTETLPTYSNLIHDI 383
OY   366 QENGFLSSLPIYIGSLMCLMILSGAADNLRAKNPFSTLCVRRLFSLIGMIGAVPLVAGF 425
DB   384 QSNALLNVPYITTSFVFGIACSLADMMLARYISILNSYKIMNTVASVBSLGLIGIT 443
OY   426 IGCDDSLVAVFITSTTGGFCSSGFSINHLDIAPSVAGILGINTFATPIGMWGP-VI 484
DB   444 VCGDWVWYTFMLAGISGSGAVYAGNQNHHIALSPRYAGTGTINSANSANICGFLAPYI 503
OY   485 AKSLPTDNTVGEWQTVFYIAAIVNFGAIFETLFAKGEVQWMA 527
DB   504 GLIINHRETLQWHLVFWLALAGLNATAGNFYILIFASAEQSWMS 546

RESULT 5
O9VDMO PRELIMINARY; PRT; 502 AA.
O9VDMO:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CG4288 PROTEIN (GH23975P).
DE GN CG4288.
OS Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta;
OC Pterygota: Neoptera: Endopterygota: Diptera: Brachyoptera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLIN=20196006; PubMed=10731133;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Genetides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S.R., Ashburner M., Henderson S.N.,

```

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.E., Aghayani A., An H.-J., Andrews-Planckoch C., Baldwin D.,
 RA Balcer R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Buritis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., Moleod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzo J.M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Glibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Y, CN BW SP;
 RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,
 RA Chamez M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC - SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
 DR EMBL: AEO03730; AAF55770.1; -;
 DR EMBL: AY060776; AAL28324.1; -;
 DR FlyBase: FBgn0038799; CG4288.
 DR InterPro: IPR003662; sub_transporter.
 DR Pfam: PF00083; sugar_tr; 1.
 KM Transmembrane.
 SO SQUENCE 502 AA; 56112 MW; 460D6FD1DE741CE2 CRC64;

Query Match 35.0%; Score 992; DB 5; Length 502;
 Best Local Similarity 42.4%; Pred. No. 7.4e-11;
 Matches 190; Conservative 93; Mismatches 149; Indels 16; Gaps 3;

QY 80 RYNALIAFGGFTVVALRVNLASVALVDWDSNTLTEDNRTSKACRPHSAPIKVNHOQTG 139
 DB 38 RYIVNLIAFGGFTVVALRVNLASVALVDWDSNTLTEDNRTSKACRPHSAPIKVNHOQTG 83
 QY 140 KRYOMDAETGWMILGSGFFGYIITQIPGVYASKIGKMLGFGILGTAVLTPTPIAAD 199
 DB 84 -DEPMDSKQGLIISFFGYIITQIPGVYASKIGKMLGFGILGTAVLTPTPIAAD 142
 QY 200 LGVGPLIVLALBGLGEGVTPPANHAMSSWAPPLERSKILLSISYAGOLGTIVISPLSG 259
 DB 143 HSLMEFLVRIIEGFEFGVFPGHAVAMRSPLEERSRMASTIAFAAGNAGYVAVMPCSG 202
 QY 260 ITCYMMWTVVFFFGTIGTFWPLIMTWLVSDDPQKHKRISHYEKEITILLSLRNOLSSOK 319

DB 203 FLATKMGSEVFFVFGTIGIWIYTWLVFKAGBELDFCSKECDYIOKTIGVYSGKHV 262
 QY 320 SVPMVPLKSLPMALVAVHAFSWMTFTLTLLPTVMKEILRPNOENGFLSLPLSG 379
 DB 263 KHPRAIFTFMPFAITASHFSEMGFTYLLTQLPSTRLTNFLNFDLCKTILASVPLAM 322
 QY 380 WLCMILSGQADNLRANKNSTLCVRRIFSLIGMIGPAVLVAAGTICDYSLAVAFLLT 439
 DB 323 GILLAVSGYLDWLOVYGWITTOVRNFCAGFAEQTVEMMLTAYL-LDPTMSVSLTI 381
 QY 440 STILGRCSSGFSINHDIPASAGILGTTNFPATIPGAVGVIAKSLPDMTVGEMOT 459
 DB 382 AVGIGARPMGSGFVFNHDLAPQHASVLMGNGFNATIPGIVSPLTIGYVVTNOTSDEMRI 441
 QY 500 VFYIAAINFGAIFFTLFAKGEYOMA 527
 DB 442 IFFLSAGIYVGVYIYFCSGDLQEMA 469
 RESULT 6
 ID 09J112 PRELIMINARY; PRT; 582 AA.
 AC 09J112;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, last annotation update)
 DE DIFFERENTIATION-ASSOCIATED NA-DEPENDENT INORGANIC PHOSPHATE
 DE COTRANSPORTER.
 GN Rat. r. norvegicus (Rat).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Mashima H., Kojima I.,
 RT "Rat DNPI.";
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF271235; AAF76223.1; -;
 SO SQUENCE 582 AA; 64575 MW; 99A14F62E685B9E9 CRC64;

Query Match 34.2%; Score 970; DB 11; Length 582;
 Best Local Similarity 40.9%; Pred. No. 5e-69;
 Matches 201; Conservative 93; Mismatches 165; Indels 32; Gaps 9;

QY 51 RNDGEBSTDTPIPLPAP---RAEAPVC-CSA-----RNALIAFFGFTIYALRVNL 101
 DB 34 KQNRRTIELTE--DGKPLEVPEKKAPLDCDCTGFLPRRTIYAIMSGLFCISFGIRCNL 91
 QY 102 SVALLVDWDSNTLTEDNRTSKACRPHSAPIKVNHOQTGKKYORDAETGWMILGSGFFGYI 161
 DB 92 GVALVDWVNNSTI---HRGKVIKEKA-----KFNMDPEYGMHRSFFWGYI 136
 QY 162 ITQIPGVYASKIGKMLGFGILGTAVLTPTPIAADLGVPGLVLRALBGLGEGVTPP 221
 DB 137 ITQIPGVYASRLAARVFGAAILTLSTLNMILPSAARVHYGCIVFIRILQGLVEGVTPP 196
 QY 222 AMHAMSSWAPPLERSKILLSISYAGOLGTIVISPLSGIICYYMMNTYVFFFGTIGTM 281
 DB 197 ACHGISKMAWAPPLERSKILLSISYAGAVIAMPPLAGLILVOYTGSSVFFYVSGGSMW 256
 QY 282 FLIMWLVSDTPQKHKRISHYEKEITILLSL---RNLSSOK---SVPMVPLKSLPMALV 316
 DB 257 YMWLVSVSPAKHPTIDEERRYIEESIGESANILGMANKRKTTPMRKFFTSMPYAYI 316
 QY 337 VAHPTWMTFTYLLTLPTVMKEILRPNOENGFLSLPLSGWLCMILSGQADNLR 396
 DB 317 VANFCSWTFYLLISQPAFEEVEFGFEISKYMLSAVPLVWTIIVPIGGQIADFLRSK 376
 QY 397 WNFSTLCVRRIFSLIGMIGPAVLVAAGTICDYSLAVAFITISTILGRCSSGFSINHL 456
 DB 377 QILSTTVRKINMGCGMEATLLLVGY-SHTRGVAISPLVAVGFSGFAISGFVFNHL 435

OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.; Gattung S.; Lee T.T.;
 RT "The sequence of C. elegans cosmid 2K54.";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RT "Direct Submission.";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U58737; AAK39396.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 586 AA; 65045 MW; 95D52B7F7FA7DD0F CRC64;

Query Match 33.4%; Score 946.5; DB 5; Length 586;
 Best Local Similarity 42.6%; Pred. No. 3.8e-67;
 Matches 193; Conservative 79; Mismatches 154; Indels 27; Gaps 6;

QY 80 RYMLATIAFGFTYVALRNLISVALVMDSTLTEDNTSKACRPHSAPIKVHNQGT 139
 DB 145 RHVVAITALLGFANIYAMRNLISIAIVEMT-SETERKVNQTL-----HVLG 190
 QY 140 KKYQVAETQGMILGSEFFYYITITQPGYVASKIGKMLGFGILCTAVLTFTPAAD 199
 DB 191 DFNWMPMNOGVVLSFFYYIYISQPLGTYLATTHGAKTIFFGAGTGTAVTTLTPPRAR 250
 QY 200 LGVGPLIVLRALGELGEGVTFRPMHAWSSWAPDLERSKLLSTISYAGAOGLTVISLPLSG 259
 DB 251 MGYGMILVFAFMFGELLEGVYYPAMHVIWMSRMAPMEQTKLATPAFSGSYGTVAAPLSA 310
 QY 260 IICYYMMTVVFEFFEGIGTFMFLMTLVSDTPQKHKRISHYEKEIISLNQSSOK 319
 DB 311 YLDEHFGWPIEFWFEGALGVIMCWVWKTVHDPEDDPKISTSE---LALLQPDAYSON 366
 QY 320 S--VPWVPIIKSLPLMAIVVAHFSYMTFTYLLTLLPTMYKEILRFNVQENGFLSLPYL 377
 DB 367 HYIVPAQAILRSKRPVNAVIVAHSAQNGIFYIMLTNLPKMLKDLAGIVNEAGLASSLPYF 426
 QY 378 GSWLMLSGQAADNLRAKNFSTLCVRRISLSLGMIGRAVPLVAAAGFTICDYS--LAV 434
 DB 427 LMGFOIITGGOLCDYLRBDHNDTLFVRKMACALGFTGCVFL---FLVMTTSNLSLLV 482
 QY 435 APLTISTTLGGFSSGSRSTNHLDIAPSYAGILGTTTPTFTPGMGOPVYAKSLTDPNTV 494
 DB 483 LFFSISIGLGCWCGFSVNHDLAPQYAGHLMATSNFTATPITGICPLLVGAIVONGTI 542
 QY 495 GEMQVEYIAAIVFGAIFETLFAKGEVONMA 527
 DB 543 GEMNVIMYIIISAVLGAALTKFADATTDLPWA 575

RESULT 10
 O9P2U7 PRELIMINARY; PRT: 560 AA.
 AC O9P2U7;
 DT 01-DEC-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, last annotation update)

DE BRAIN-SPECIFIC NA-DEPENDENT INORGANIC PHOSPHATE COTRANSPORTER.
 GN BNPI.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RA Aihara Y., Mashima H., Onda H., Hisano S., Kasuya H., Hori T.,
 RA Yamada S., Tomura H., Yamada Y., Inoue I., Kojima I., Takeda J.;
 RT "Molecular cloning of a novel brain-type Na+-dependent inorganic
 phosphate cotransporter.";
 RL J. Neurochem. 0:0-0(2000).
 DR EMBL: AB032436; BAA92875.1; -.
 SQ SEQUENCE 560 AA; 61613 MW; C8BDAPB34B6E45B6 CRC64;

Query Match 33.2%; Score 941; DB 4; Length 560;
 Best Local Similarity 39.5%; Pred. No. 9.8e-67;
 Matches 193; Conservative 93; Mismatches 162; Indels 40; Gaps 8;

QY 53 DGEESDRT---PLLP-----GAPRAEAPVCCSARKNLALAFGFFIYALAVNLSVAL 105
 DB 38 DGRPVTTQTDRDPVVDCTCFGLPR-----RYIIAISMGLGFCISFGIRCNLGVAI 87
 QY 106 VDWVDSNTLTEDNRTSKACEHSAPIKVHNQGTQKQWMAEFGWTLGSEFFYYITQI 165
 DB 88 VSMVNNSTT-----HRGHVYVQAKQ---FSMPDETGVGLHGSFFMGVITQI 132
 QY 166 PGGVYASKIGKMLGFGILGTAVLTFTPIADLGVPLIVLRALGELGEGVTFPAMA 225
 DB 133 PGCFIQCKFANVFPGRPAIVATSTLNLIPSAKVHGCYIFRILQGLGEGVTPACRG 192
 QY 226 MMSWAPDLERSKLLSTISYAGAOGLTVISLPSGIICYYMMTVVFEFFEGIGTFMFLW 285
 DB 193 IWSKMAPDLERSLATTAFFCGSYAGAVAMPAGVLYOVYSGMSVFFVYSGIFMYLFW 252
 QY 286 IWLVSQDPQKHKRISHYEKEI-----LSLNQSSOKSVMPVPIKSLPLMAIVVAH 339
 DB 253 LTVYESPALHPSISEERKVTIEDAIGESAKLKNPL-TKSTYWRREFTSMYPALIVAN 311
 QY 340 FSYNMTFTYLLTLLPTMYKEILRFNVQENGFLSLPSLGMILSGQAADNLRAKNF 399
 DB 312 FCRSMFTYLLISQPAFEEVFEGFELISKVGLVALPHLVITIIIPGIGQADLRLSRIRM 371
 QY 400 STLTVRRISLIMIGPAVFLVAAAGFTICDYSIAVAFITSTTLGGFSSGSRSTNHLDIA 459
 DB 372 STTNVRLKMGCGFGMEATLLLVGY-SHSGVYASIFLVAVGFSGRFASGFVNHLDIA 430
 QY 460 PSYAGILGTTNTPFATIPGAVGVIAKSLRPMQVGMQVVFYIAAIVFGAIFETLPA 519
 DB 431 PRYASTILMGISNGVTLGKWCPIIVGAMTKHKTREMOYVPLIASLVHGVIFYGEVA 490
 QY 520 KGEVONMA 527
 DB 491 SGEKQRPWA 498

RESULT 11
 O62634 PRELIMINARY; PRT: 560 AA.
 AC O62634;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE BRAIN SPECIFIC NA+-DEPENDENT INORGANIC PHOSPHATE COTRANSPORTER.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.

```

RX MEDLINE-94261635; PubMed-8202535;
RA N1 B., Rostock P.R., Nadi N.S., Paul S.M.;
RT "Cloning and expression of a cDNA encoding a brain-specific Na(+)-
RT dependent inorganic phosphate cotransporter";
RL Proc. Natl. Acad. Sci. U.S.A. 91:5607-5611(1994).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL; 007609; AAA19646.1; -.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR Transmembrane.
SQ SEQUENCE 560 AA; 61665 MW; F686889F606B8305 CRC64;

Query Match 33.0%; Score 936; DB 11; Length 560;
Best Local Similarity 39.0%; Pred. No. 2.5e-66;
Matches 190; Conservative 91; Mismatches 168; Indels 38; Gaps 7;

QY 53 DDEESTDRT---PLLP---GAPRAEAPVCCSARYNLAILEFFGFYVALRVNLVAL 105
DB 38 DGRPVYTHRRDPVYVDCVCFGLPR-----RYIIMSGLCISFGIRCNLGVAI 87
QY 106 VDMVDSNTLLEDRKSKACPEHSAPIKYAHNQTGKRYQMDAETGQILDSFFYGIITQI 165
DB 88 VSMVNSST-----HRGHHVYVOKAQ---FNMDPEVYGLIHSGFFGMYITQI 132
QY 166 PGYVASKIGKMLGFGILGAVLTFPIPIADLGVPLIVRALEGEGEYTPPAMA 225
DB 133 PGGFQICAAANRVGFALVATSTLMLIPSAARHYGCVIFVRILOGIVEGTYTACHG 192
QY 226 MMSWAPPLERSKLLISYAGAOLGTIVSLPGIICYYMMNTYVYFPGTIGIEFWLLM 285
DB 193 INSKAPPLERSLATTATCGSYAGAVVAMPLAGVLQVSGMSVYVYVYSGGIEFWYLM 252
QY 286 IMLVSTPOKHKRISYEVKYLISRLNLSQSKY-----FWVPLKSLPMAIVAAH 340
DB 253 LLYVSEPALHPSISEEKRIYEDAGESAKLMPYTKNTPRRPFTMPYVAIIVANF 312
QY 341 SYNWFYTLTLPTMYKELIRFNVQENGLSLPYLGSLMCLMISGOADNLRAKWNS 400
DB 313 CSWTFYLLILISOPAFEEVFGFEISKVGLVSLPLHWTIIVYIGQIADFLRSRHNS 372
QY 401 TLCVRIFSLIGMIGPAFLVAGFICDYSLAVAFLLISTTLGFCSSGFSINHLDIAP 460
DB 373 TTNVRKLMCGEGMEATLLLVGY-SHSKGYAISFLVAVGSGRAISGFENVHLDIAP 431
QY 461 STAGILIGITNFATIPGAVGPIAKSLTPDNTVGEMQVYFIAAINFGALFTFLPK 520
DB 432 RYASILMGISNGVGLSGMVCPLIVGAMTKHKTREMOYVFILASLVHGYVIFGVFAS 491
QY 521 GEYQWMA 527
DB 492 GEGQPPA 498

RESULT 12
Q9V7S5 PRELIMINARY; PRT: 529 AA.
AC Q9V7S5; Q9V7S6; O61364;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE INORGANIC PHOSPHATE COTRANSPORTER (PICOOT PROTEIN) (CG8098
DE PROTEIN).
GN PICOOT OR CG8098.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RC STRAIN-BERKELEY;

RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celnicher S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.H.C., Yandell M.D., Zhang Q., Chen L.X.,
RA Burton R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Abrell J.F., Abmayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Ewanigista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kensington J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacible J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,
RA Reibert K., Remington I., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spierias R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasthman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.F., Zaveri J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE OF 449-529 FROM N.A.
RC STRAIN-CANTON-S;
RA Da Lage J.-L.;
RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS; LONG ISOFORM (SHOWN HERE) AND
CC SHORT ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL; AE003806; AAF57968.1; -.
DR EMBL; AE003806; AAF57969.1; -.
DR EMBL; AF022713; AAD09148.1; -.
DR FlyBase; FBgn0024315; Picoot.
DR InterPro: IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
KW Alternative splicing; Transport; Transmembrane.
FT TRANSMEM 37 57
FT TRANSMEM 148 168
FT TRANSMEM 202 222
FT TRANSMEM 232 252
FT TRANSMEM 338 358
FT TRANSMEM 429 449
FT TRANSMEM 466 486
FT TRANSMEM 498 529
FT VARSPLIC 1 35
SQ SEQUENCE 529 AA; 58372 MW; 07B89A52D5081EFB CRC64;

Query Match 32.7%; Score 928.5; DB 5; Length 529;
Best Local Similarity 41.1%; Pred. No. 9e-66;
Matches 190; Conservative 81; Mismatches 176; Indels 15; Gaps 8;

QY 76 CCSARYNLAILEFFGFYVALRVNLVDMVDSNTLLEDRKSKACPEHSAPIK 132
DB 36 CEAIRYFVEMFLGMANAYVMRTNMSVAIVAV-NHTAIKSGEAEYDDECGDRDIP- 93

```

| | | | |
|----|-----|---|-----|
| Oy | 133 | VHNNOTGKRYOMDAETQWMLIGSFYGIITQIRGCVYASKIGSKMLLFGGILGTAVLT | 192 |
| Db | 94 | -DDSDG:-EFAMSSALOYILISSEFYGVITQIPFGILAKYGSIRLQGLMINSYFAR | 151 |
| Oy | 193 | FTPIA-PLGAGPLILVLEALGLEEGYTFPMHMMSSNAPRLERSKLLSTISVAGDLGT | 251 |
| Db | 132 | LVPVAARGGVWGLCAVAFIOGLDEGRVPTNHMLAKMIPRNRSMGAAYVAGAOFGT | 211 |
| Oy | 232 | VISLPLSGIICY--NMNVTVEFFEGTIGIFWELMIWLSDTPOKHRIKSHYEKEYIL | 308 |
| Db | 212 | IISMPLSGLIAEYGDGGMPSIFVFEYGVIGVWSIAFLIVHEDPSSHPTIDEREKKYIN | 271 |
| Oy | 309 | SSL-RNQLSSOKSVWPVWILKSLEPLMAIVAHSEYNNTFYTLTLFTYKELLREVQ | 367 |
| Db | 272 | DSLMTGDVVKSPPIEFKAIRKISLFEYALIFPAHMHNNGYEFLMTELPYRKQVLRSELS | 331 |
| Oy | 368 | NGFLSLPYLGSWICMLISGOAANLRKKNWFSYLCYRIFSLIGMIPVAVYAGETG | 427 |
| Db | 332 | NCGLSSLYTLAMMLEFSMTISYVADMMTSSKRFSTIATRKLLNSIGCYGPGVALIAMS | 351 |
| Oy | 438 | CDYSLAVALFIISTTLGGFCCSGFSINHLDIAPSYAGILGINTFAETPIGMMGVPIAKS | 487 |
| Db | 392 | CDRAITLAILITIGVLANGIYSGFKINHLLDTPRAGFMSITNCSANLGLAPIAGH | 451 |
| Oy | 488 | LTPDNT---VGEKQTVPIIAAINVEGALPITLFAKEVQNW | 526 |
| Db | 452 | LISPSKPMGQOIVFEIIFALFVYIIIGTFENIGSEBERYW | 493 |

| Query Match | 32.3% | Score 917 | DB 5 | Length 483 |
|-----------------------|--|-------------------|-----------|------------|
| Best Local Similarity | 41.7% | Pred. No. 6,7e-65 | | |
| Matches 188 | Conservative 82 | Mismatches 167 | Indels 14 | Gaps |
| QY 86 | LAFFGFFIVYALRNLSVALVDWDSNTTLE--DNRTSKACPEHSAPIKVHNQTKKTY 143 | | | |
| DB 1 | MLFLGMANAYVNRKMSVAIYAWY-NHTALKSGEEETDEDCGRDPI--DSDQD-EFP 56 | | | |
| QY 144 | WDAETQGMILGSFFGYVITIQIPGGYVASKIGCKMILGFGILCTAVLLTFTPIAA-DLGV 202 | | | |
| DB 57 | WNAALQGIITLSSFFGYVITQIPGILAKKYSPLRFYGMILNSYFAFLVPAAREGV 116 | | | |

| | | | |
|----|-----|--|-----|
| QY | 203 | GPVLVRLBEGIGCVTFPRAMHAMMSNAPPLESKLSTSYAGAOGLTVISLPSLITC | 262 |
| Db | 117 | WGLCAVREIIOGIGEDPVPCTHAMLAKWIPENESRMCAGVAACPGCTIIMSGLLA | 176 |
| QY | 263 | YX---MMNTVYFFPGTIGTFWFLMTIMVSDTPQKHKRISHYKEVILSSL-RNQSSQ | 318 |
| Db | 177 | EYGEDGGMPSIFYVGVIGVWASIAFLFLEYEDSTPHKIDERRKTYINSLMGTDIKS | 236 |
| QY | 319 | KSVPPVPLIKSLPMAIVAAHFNSMTEFVTLTLPLPYKMEILFENOQMSFLSLPYLG | 378 |
| Db | 237 | PIIPKSLVKSIPFAILFAIMGNINCYETIMTELPYMKQVLRFLSKNGLSLPYLA | 296 |
| QY | 379 | SWLMLISGQADMLRAKNWSTLCVRRIFSLIGMIGPAVFLVAAGFIGDYSLAAVFLT | 438 |
| Db | 297 | MWLLSMFISYIADWMWISSKRFSLPATRKLIINSIGQYOGBALIAASYTGCORALTIALT | 356 |
| QY | 439 | ISTTLGGCCSSGFSINHLDIAPSTAGILGTTNFAIIPGNVGVIAKSLTPDNT---VG | 495 |
| Db | 357 | IGVGINGGIYSGFKINHLDLPRPAGFLMSTITNCSANLAGLAPIAAGNLISDPSKPVWG | 416 |
| QY | 496 | EMQPFVYFAALINVGAFIEFLFLFAKGEVQNM | 526 |
| Db | 417 | OMQVIEFAAEVYIICTGTETINFSGGEQFM | 447 |

| | | | |
|----------|---|--------------|--------------|
| 096LHI | 14 | | |
| ID | 096LHI | PRELIMINARY; | PRT; 466 AA. |
| AC | 096LHI; | | |
| DT | 01-DEC-2001 (TEMBREL, 19, Created) | | |
| DT | 01-DEC-2001 (TEMBREL, 19, last sequence update) | | |
| DT | 01-DEC-2001 (TEMBREL, 19, last annotation update) | | |
| DE | D1139621.1.2 (SODIUM PHOSPHATE SOLUTE CARRIER FAMILY 17 MEMBER 2 (ISOPORM 2)) (FRAGMENT). | | |
| GN | SLC17A2. | | |
| OS | Homo sapiens (Human). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. | | |
| OX | NCBI_Taxid=9606; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RA | Peck A.; | | |
| RL | Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases. | | |
| DR | EMBL, AL138726; CAC69102.1; -. | | |
| FT | NON_TER | 1 | |
| SEQUENCE | 466 AA; 50664 MW; F64CCE5FC89E5FC3D CRC64; | | |

| | | | |
|----------|---------|-----------|-------------------------|
| SEQUENCE | 466 AA; | 50664 MW; | F64CE5FC89E5FC3D CRC64; |
|----------|---------|-----------|-------------------------|

| | | | | |
|-------------|--------|------------|-------|-------------|
| Query Match | 31.68; | Score 895; | DB 4; | Length 466; |
|-------------|--------|------------|-------|-------------|

Best Local Similarity 39.88; Pred. No. 3.6e-63,

| | | | | | | | | | |
|---------|------|--------------|-----|------------|------|--------|-----|------|----|
| Matches | 183; | Conservative | 92; | Mismatches | 173; | Indels | 12; | Gaps | 5; |
|---------|------|--------------|-----|------------|------|--------|-----|------|----|

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 | 191 | 192 | 193 | 194 | 195 | 196 | 197 | 198 | 199 | 200 | 201 | 202 | 203 | 204 | 205 | 206 | 207 | 208 | 209 | 210 | 211 | 212 | 213 | 214 | 215 | 216 | 217 | 218 | 219 | 220 | 221 | 222 | 223 | 224 | 225 | 226 | 227 | 228 | 229 | 230 | 231 | 232 | 233 | 234 | 235 | 236 | 237 | 238 | 239 | 240 | 241 | 242 | 243 | 244 | 245 | 246 | 247 | 248 | 249 | 250 | 251 | 252 | 253 | 254 | 255 | 256 | 257 | 258 | 259 | 260 | 261 | 262 | 263 | 264 | 265 | 266 | 267 | 268 | 269 | 270 | 271 | 272 | 273 | 274 | 275 | 276 | 277 | 278 | 279 | 280 | 281 | 282 | 283 | 284 | 285 | 286 | 287 | 288 | 289 | 290 | 291 | 292 | 293 | 294 | 295 | 296 | 297 | 298 | 299 | 300 | 301 | 302 | 303 | 304 | 305 | 306 | 307 | 308 | 309 | 310 | 311 | 312 | 313 | 314 | 315 | 316 | 317 | 318 | 319 | 320 | 321 | 322 | 323 | 324 | 325 | 326 | 327 | 328 | 329 | 330 | 331 | 332 | 333 | 334 | 335 | 336 | 337 | 338 | 339 | 340 | 341 | 342 | 343 | 344 | 345 | 346 | 347 | 348 | 349 | 350 | 351 | 352 | 353 | 354 | 355 | 356 | 357 | 358 | 359 | 360 | 361 | 362 | 363 | 364 | 365 | 366 | 367 | 368 | 369 | 370 | 371 | 372 | 373 | 374 | 375 | 376 | 377 | 378 | 379 | 380 | 381 | 382 | 383 | 384 | 385 | 386 | 387 | 388 | 389 | 390 | 391 | 392 | 393 | 394 | 395 | 396 | 397 | 398 | 399 | 400 | 401 | 402 | 403 | 404 | 405 | 406 | 407 | 408 | 409 | 410 | 411 | 412 | 413 | 414 | 415 | 416 | 417 | 418 | 419 | 420 | 421 | 422 | 423 | 424 | 425 | 426 | 427 | 428 | 429 | 430 | 431 | 432 | 433 | 434 | 435 | 436 | 437 | 438 | 439 | 440 | 441 | 442 | 443 | 444 | 445 | 446 | 447 | 448 | 449 | 450 | 451 | 452 | 453 | 454 | 455 | 456 | 457 | 458 | 459 | 460 | 461 | 462 | 463 | 464 | 465 | 466 | 467 | 468 | 469 | 470 | 471 | 472 | 473 | 474 | 475 | 476 | 477 | 478 | 479 | 480 | 481 | 482 | 483 | 484 | 485 | 486 | 487 | 488 | 489 | 490 | 491 | 492 | 493 | 494 | 495 | 496 | 497 | 498 | 499 | 500 | 501 | 502 | 503 | 504 | 505 | 506 | 507 | 508 | 509 | 510 | 511 | 512 | 513 | 514 | 515 | 516 | 517 | 518 | 519 | 520 | 521 | 522 | 523 | 52 |
|--|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
|--|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|

QY 77 CSARNALIAFFGFIVYALRVNLSVALVMDVSNITLEDNRKSCAPE-----HSA 129

Db 2 CSLRYGLALIMHESNFTMITQRVSLIAIAMV - -NTT-QQQLSNASTEGPVADAFNNS 58

QY 130 PIKVHNOT-GKKYQWDAETQWILGSFFYGYITITPIPGYVASKIGCKMLLGEIIGTA 18

```

Db      59 SISKEDTKASYQWSPETOGIIFSSINGLITLIPSGYLAGIFGAKKMLGAGLITS 118

```

OV 189 VLTFTPIADLGVGPIVLRALEGI GEGVTEPAMHAMSSWAPPIERSKLISTSYAGAO 248

[illegible]

DD 119 LLI LFI PLAADEGVLLVIMVKI VQGMQGMWIGUFI IMAKWAPEKSKLIIJAGSGSA 1/8

QY 249 LGTVISLPISGIIICYMNWTVEFFFGTIGIFWELLMIWLVSDDTPQKKHKKRISHYEKEYL 308

Db 179 FGSFIILCVGGLISQALSWPFIFYIFGSGVCVCLLMTFTVIYDDPMHNPICISVREKEHL 238

QY 309 SSLRQLOSSQ-KSVPWPILKSLPLWAIWAHFSYNTFTYLLTLLPTYMKEILRENOE 367

```
Db      239  SSLAQDPSSPGRVPIKAMVTCLPLMAIFGFFSHFWLCTIILYLPITYISTLLHVNIRD 2986
```



```

OY 368 NCFSLSLPYLGSLWCMILSGQADNLRAKWNSTLCVRRESLIMIGPAVFLVAGFTG 427
      :|||||: : ||| | : : : : : : : : : : : : : : : : : : : : : : :
Db 299 SCVLESLPEFIAASCITIGGQIADFLIRNLRLITVRKLESLDILPSICAVLPAFVA 358
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 428 CDYSLAVAFILITSTLGGCSGFSINHLDIAPSYAGILITNTFATIPGWPVIKAS 487
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 SSTVITITLILIPSTSNICDSGFIITNLDIAPRYASFMLGSRGFLAGIISTATGEF 418
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 488 LPPDNTVGEWQTVFYIAAIVNFGAIFFTLFAKGEVQNMMA 527
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 419 LISQDFESGWRNVFFLSAANMFGLVFLITPGQALQDMA 458
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15

O9Y2C5

PRELIMINARY: PRT: 497 AA.

ID O9Y2C5

AC O9Y2C5

DT 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE NA/PO4 COTRANSPORTER HOMOLOG.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI_taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99253143; PubMed=10319585;

RA Shihui A., Tsunoda T., Seki N., Suzuki Y., Sugane K., Sugano S.;

RT "Isolation and chromosomal mapping of a novel human gene showing

RT homology to Na+/PO4 cotransporter.";

RL J. Hum. Genet. 44:190-192(1999).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.

DR EMBL; AB020527; BAA76663.1; -

DR InterPro; IPR003662; sub_transporter.

DR Pfam; PF00083; sugar_tr; 1.

DR KX

DR KW

SQ SEQUENCE 497 AA; 54055 MW; 5A5616AA52D4990D CRC64;

Query Match 30.9%; Score 876; DB 4; Length 497;

Best Local Similarity 37.8%; Pred. No. 1.3e-61;

Matches 176; Conservative 98; Mismatches 182; Indels 10; Gaps 3;

```

OY 77 CSARNYLAIIAFEGFIYALRVNLSVALVDMVDSNTLIEDNRTSKACPEHSAPIKVHN 136
      ||| :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 33 CSVRGLALIIQLCNFSITYQOMNLISAIIPAV--NNTAPPSPQPNASTERPSTDQGYWN 90
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 137 QTGR-----YQDAETQWILGSFFGYITIQPGYVASKIGKMLGFGILGTAV 189
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 91 ETLKFKAMAPRAYDMSPEIQIILSLNNGSLAPISGYVAGITGAKYVAGAGLFISSF 150
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 190 LTLFTPIADLGVPLIVLRALEGLGEGVTFPAMHAMSSWAPPLERSKLSISYAGQL 249
      ||| :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 151 LTLFTPIAANAGVALLIYLRIVQIAQVVLGQYSIMWKAPPLERSQLTIAGSGSML 210
      ||| :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 250 GTVVISLPISGIICTYMMNTYVYFEGTIGIFWFLIMWLVSPTPOKHKRISHYEKEYILS 309
      ||| :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 211 GSEIVLLAGGLCQITIGWYVYIFGIGACCPLMFPLIYDDPVNHPFISAGEKRYIVC 270
      ||| :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 310 SLRNO-LSSQKSVWPVPIKSLPLMAIVANFYSYNTFTYLLTLPTMKELIRFNVQEN 368
      ||| :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 271 SLAQDQCPGWSLPIRAMIKSLPLMAIIVSYCEYEWLFTTMAVPTIYISSVLQANLDS 330
      ||| :||| : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 369 GFLSLPYLGSLWCMILSGQADNLRAKWNSTLCVRRESLIMIGPAVFLVAGFTGC 428
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 331 GILSLPEYVVGICITILGLADFLIRNLRLITVRKLESLDILPSICAVLPAFVA 390
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 429 DYSILAVAFILITSTLGGCSGFSINHLDIAPSYAGILITNTFATIPGWPVIKASL 488
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 SHSMITFLVLSAISFCESGALVNFIDIAVRYTGFLKGLQVFAHIAAISPTAAGFF 450
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

OY 489 TPDNTVGEWQTVFYIAAIVNFGAIFFTLFAKGEVQNMALNDHGH 534
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 451 ISQDFESGWRNVFFLSAANMFGLVFLITPGQALQDMA 496
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: July 15, 2002, 08:25:02
 Job time: 167 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 08:22:15 ; Search time 56.81 seconds
(without alignments)
1047.977 Million cell updates/sec

Title: US-09-776-865-2
Perfect score: 2836
Sequence: 1 MAAGAMTPRPVQPARPGF.....LFAKGEVQNALNDHGHRRH 536

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_032802:*

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-----------------------------|
| 1 | 2836 | 100.0 | 536 | 21 | AAV45089 Human GBS toxin re |
| 2 | 2836 | 100.0 | 536 | 22 | AAE06518 Human group B beta |
| 3 | 2621 | 92.4 | 495 | 21 | AAV45087 Patial human GBS |
| 4 | 2621 | 92.4 | 495 | 22 | AAAB6697 Human ASF. Homo s |
| 5 | 2614 | 92.2 | 495 | 22 | AAAM3895 Human polypeptide |
| 6 | 2329 | 82.1 | 495 | 21 | AAV45088 Sheep GBS toxin re |
| 7 | 2329 | 82.1 | 495 | 22 | AAE06519 Sheep group B beta |
| 8 | 1698 | 59.9 | 495 | 21 | AAV45090 Human/Sheep consen |
| 9 | 1679 | 59.2 | 314 | 22 | AAAG6538 Human sodium depen |
| 10 | 1614 | 56.9 | 309 | 22 | AAW40745 Human polypeptide |
| 11 | 1450 | 51.1 | 277 | 22 | AAW93914 Human polypeptide |

| | | | | | |
|----|-------|------|-----|----|-------------------------------|
| 12 | 996.5 | 35.1 | 559 | 22 | ABB58701 Drosophila melanog |
| 13 | 992 | 35.0 | 502 | 22 | ABB60525 Drosophila melanog |
| 14 | 992 | 35.0 | 502 | 22 | ABB65873 Drosophila melanog |
| 15 | 959.5 | 33.8 | 582 | 22 | AAW92973 Human protein S60 |
| 16 | 941 | 33.2 | 567 | 22 | AAO13870 Human polypeptide |
| 17 | 940 | 33.1 | 560 | 17 | AAW05148 Human brain sodium |
| 18 | 940 | 33.1 | 560 | 19 | AAW70500 Human sodium-lithi |
| 19 | 939 | 33.1 | 194 | 22 | AAW25685 Human protein sequ |
| 20 | 928.5 | 32.7 | 516 | 22 | ABB67013 Drosophila melanog |
| 21 | 928.5 | 32.7 | 529 | 22 | ABB63684 Drosophila melanog |
| 22 | 856 | 30.2 | 479 | 22 | ABB62841 Drosophila melanog |
| 23 | 852.5 | 30.1 | 496 | 22 | ABB61407 Drosophila melanog |
| 24 | 838.5 | 29.6 | 481 | 22 | ABB64204 Drosophila melanog |
| 25 | 821.5 | 29.0 | 512 | 22 | ABB60925 Drosophila melanog |
| 26 | 811 | 28.6 | 493 | 22 | ABB59580 Drosophila melanog |
| 27 | 808 | 28.5 | 576 | 20 | AAW86523 Eac-4 protein amn |
| 28 | 798 | 28.1 | 465 | 22 | ABB64710 Drosophila melanog |
| 29 | 782.5 | 27.6 | 497 | 22 | ABB70142 Drosophila melanog |
| 30 | 762 | 26.9 | 475 | 22 | ABB67155 Drosophila melanog |
| 31 | 762 | 26.9 | 491 | 22 | ABB70144 Drosophila melanog |
| 32 | 762 | 26.9 | 560 | 22 | ABB59401 Drosophila melanog |
| 33 | 756.5 | 26.7 | 462 | 22 | ABB70143 Drosophila melanog |
| 34 | 755 | 26.6 | 343 | 22 | ABG12999 Novel human diagno |
| 35 | 747 | 26.3 | 512 | 21 | AAAG29915 Arabidopsis thailia |
| 36 | 747 | 26.3 | 519 | 21 | AAAG29914 Arabidopsis thailia |
| 37 | 731 | 25.8 | 395 | 21 | AAAG29916 Arabidopsis thailia |
| 38 | 729 | 25.7 | 449 | 22 | ABB58659 Drosophila melanog |
| 39 | 708 | 25.0 | 256 | 22 | ABG02032 Novel human diagno |
| 40 | 696.5 | 24.6 | 424 | 22 | ABB64935 Drosophila melanog |
| 41 | 686 | 24.2 | 436 | 19 | AAW78919 Human haemochromat |
| 42 | 653 | 23.0 | 425 | 22 | AAAG81307 Human AFP protein |
| 43 | 643 | 22.7 | 194 | 22 | ABB64689 Drosophila melanog |
| 44 | 615.5 | 21.7 | 401 | 19 | AAW78920 Human haemochromat |
| 45 | 615.5 | 21.7 | 401 | 19 | AAW69971 Human sodium-depen |

ALIGNMENTS

| | |
|----------|---|
| RESULT 1 | |
| ID | AAV45089 standard; Protein: 536 AA. |
| XX | |
| AC | AAV45089; |
| XX | |
| XX | 31-MAY-2000 (first entry) |
| XX | |
| DE | Human GBS toxin receptor (HP59). |
| XX | |
| DT | |
| XX | |
| XX | Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59; |
| KW | pathological vascularisation; Cancer metastases; angiogenesis; |
| KW | neovascularisation; reperfusion injury; scarring; keloid; |
| KW | chronic inflammation; rheumatoid arthritis; psoriasis; neural injury; |
| KW | endothelial cell proliferation; antibacterial; anticancer; |
| KW | anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic. |
| OS | |
| XX | Homo sapiens. |
| XX | |
| EH | |
| FT | Key |
| FT | Location/Qualifiers |
| FT | /note= "Putative amidation site" |
| FT | 138..141 |
| FT | Modified-site |
| FT | /note= "Putative amidation site" |
| FT | 100..103 |
| FT | Modified-site |
| FT | /note= "Asn is N-glycosylated" |
| FT | 112..115 |
| FT | Modified-site |
| FT | /note= "Asn is N-glycosylated" |
| FT | 118..121 |
| FT | Modified-site |
| FT | /note= "Asn is N-glycosylated" |
| FT | 136..139 |
| FT | Modified-site |
| FT | /note= "Asn is N-glycosylated" |
| FT | 266..269 |
| FT | Modified-site |
| FT | /note= "Asn is N-glycosylated" |

| | | |
|----|--------------------------|---|
| FT | Modified-site | 343..346 /note= "Asn is N-glycosylated" |
| ET | Modified-site | 398..401 /note= "Asn is N-glycosylated" |
| ET | Modified-site | 297..300 /label= CAMP_phospho_site /note= "Putative phosphorylation site" |
| ET | Modified-site | 113..116 /label= CR2_phospho_site /note= "Putative phosphorylation site" |
| ET | Modified-site | 114..117 /label= CR2_phospho_site /note= "Putative phosphorylation site" |
| ET | Modified-site | 300..303 /label= CR2_phospho_site /note= "Putative phosphorylation site" |
| ET | Modified-site | 493..496 /label= CR2_phospho_site /note= "Putative phosphorylation site" |
| ET | Modified-site | 66..71 /note= "Putative myristylation site" |
| ET | Modified-site | 167..172 /note= "Putative myristylation site" |
| ET | Modified-site | 183..188 /note= "Putative myristylation site" |
| ET | Modified-site | 213..218 /note= "Putative myristylation site" |
| ET | Modified-site | 246..251 /note= "Putative myristylation site" |
| ET | Modified-site | 250..255 /note= "Putative myristylation site" |
| ET | Modified-site | 378..383 /note= "Putative myristylation site" |
| ET | Modified-site | 427..432 /note= "Putative myristylation site" |
| ET | Modified-site | 444..449 /note= "Putative myristylation site" |
| ET | Modified-site | 464..469 /note= "Putative myristylation site" |
| ET | Modified-site | 468..473 /note= "Putative myristylation site" |
| ET | Modified-site | 23..25 /label= PKC-phospho_site /note= "Putative phosphorylation site" |
| ET | Modified-site | 58..60 /label= PKC-phospho_site /note= "Putative phosphorylation site" |
| ET | Modified-site | 78..80 /label= PKC-phospho_site /note= "Putative phosphorylation site" |
| ET | Modified-site | 120..122 /label= PKC-phospho_site /note= "Putative phosphorylation site" |
| ET | Modified-site | 138..140 /label= PKC-phospho_site /note= "Putative phosphorylation site" |
| ET | Modified-site | 310..312 /label= PKC-phospho_site /note= "Putative phosphorylation site" |
| ET | Modified-site | 317..320 /label= PKC-phospho_site /note= "Putative phosphorylation site" |
| XX | WO200005375-A1. | |
| XX | 03-FEB-2000. | |
| XX | 22-JUL-1999; | 99WO-US16676. |
| XX | 22-JUL-1998; | 98US-0093843. |
| XX | (UYVA-) UNIV VANDERBILT. | |

PI Hellergqvist CG, Fu C;
 DR WPI: 2000-205377/18.
 DR N-PSDB: AA250879.
 XX
 XX
 PT New polynucleotide encoding mammalian receptor for streptococcus toxin,
 PT useful for diagnosis and treatment of, e.g. pneumonia in neonates -
 PS
 PS Claim 10; Page 93-95; 109pp: English.
 XX
 XX The present sequence is partial human GBS (group B beta-haemolytic
 CC streptococci) toxin receptor (HP55). GBS toxin receptor is an integral
 CC protein with seven transmembrane domains. Expression vectors comprising
 CC the coding region can be transformed into host cells to express GBS
 CC toxin receptor and its fragments. Detecting the receptor in tissues is
 CC used to diagnose pathological vascularisation, e.g. for detecting cancer
 CC metastases. GBS toxin receptors are useful for treating conditions
 CC associated with pathological angiogenesis or neovascularisation
 CC (specifically cancer, reperfusion injury, scarring during wound healing,
 CC keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
 CC neural injury), and to raise specific antibodies used for treating early
 CC onset disease. Inhibitors of this receptor are useful for treating
 CC pathological or hypoxia-induced endothelial cell proliferation and
 CC migration.
 XX
 XX Sequence 536 AA:
 SQ

| Query Match | 100.0% | Score 2836 | DB 21 | Length 536 |
|-----------------------|---|--------------------|--------------|------------|
| Best Local Similarity | 100.0% | Pred. No. 1,3e-283 | | |
| Matches 536 | Conservative | 0 | Mismatches 0 | Gaps 0 |
| QY 1 | MAAGAMPTRPRPOAPRPGFGLSGRRSLLCQVASTPAHVGMKSPVPRDARNDGESTR | 60 | | |
| Db 1 | maagamtprprpoaprpqpgfglsgrtsllcqvaasphavgmrsprdlarndgeestr | 60 | | |
| QY 61 | TPLLGARRAEAPPCCSARNYLLALFFEGFTIVALKRNLSVALYDMYDSTLTLEDNRT | 120 | | |
| Db 61 | tplpgarpraeapccsarynlalaffgftivalkrnlsvalydmvdstltlednrt | 120 | | |
| QY 121 | SKACEHSAPRIKVHHNNGKRYQMDAETQGMILGSPFYGYITTOJPGCVASKIGAKML | 180 | | |
| Db 121 | skacehsaprlkvhhnngqkkyqwdaeqgvallgsftfygylttqjpgvyasklqkml | 180 | | |
| QY 181 | GFGLTGAVALTFPTPIADLVGVLIVLRALGELGBCVTFPAMHAMSSWAPPLERSKLL | 240 | | |
| Db 181 | gfglgtgvaltfltpriaadlvgypllvtraleglgegytffpamhamsswapplerskll | 240 | | |
| QY 241 | SISYGAOLGWTISPLSLGIIICYNNMTVPVFEFGTIGFWFLMWLMLSDPPOKHKRIS | 300 | | |
| Db 241 | sisygaolgtvtslplsllgiiwymwtyvifftgtlglfwllwlvysdcpqkhkris | 300 | | |
| QY 301 | HYEKYEIILSLRNQLSSOKSVPWVPLNLSPLMAIVVAHFSSYNNMFTYLLTLPTYMKEI | 360 | | |
| Db 301 | hyekyeiilslrnqssqsvpwvplkelpwaivvahfsynwtfylltlllptymkel | 360 | | |
| QY 361 | LRFNQENGLSLSLPYLGSWTCMIISGQAADLRKKNMNSTCIVRRIFSLGIMIGAVVL | 420 | | |
| Db 361 | lrfnvqengflsslpylgswlcmllsgqaadlraknmnstlcivrrifsligmigavvll | 420 | | |
| QY 421 | VAAGFICGDYSILAVFLLITSTLGGFCSSGSGFINHLDIAPSYAGILLGTTFNFPATPGWY | 480 | | |
| Db 421 | vaagfigcdysilavafllitstlggfcssgsgfinhldiapsyagillgttnftatpgnv | 480 | | |
| QY 481 | GPVIAKSLTPDNTVGEMQTVYIAAIVNFGAIFPTLPAKGEVQWALNDHHGHRH | 536 | | |
| Db 481 | gpvleaksltpdntvgewqtvfyiaaainvfgaifftllfakgevgwalnhdhghrh | 536 | | |

```

AC AAE06518;
XX
DT 16-OCT-2001 (first entry)
XX
DE Human group B beta-haemolytic Streptococci toxin receptor (HP59) protein.
XX
KW Human: group B beta-haemolytic Streptococci toxin receptor; GBS; HP59;
KW cytostatic; vulnary; antiatherosclerotic; osteopathic; vasotropic;
KW prevention; attenuation; pathoangiogenic condition; cancer; scar;
KW wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;
KW keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;
KW vaccine.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH Region 8..28
FH FT /label= Hab3_immunogenic_peptide
FH FT /label= Hab1_immunogenic_peptide
FH FT /label= Hab4_immunogenic_peptide
FH FT /label= Hab2_immunogenic_peptide
FH Region 112..125
FH FT /label= Hab2_immunogenic_peptide
XX
PN WO200156598-A2.
XX
PD 09-AUG-2001.
XX
PF 02-FEB-2001; 2001WO-US03662.
XX
PR 02-FEB-2000; 2000US-0179870.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
PI Helieryist CG;
XX
DR WPI: 2001-488844/53.
XX
DR N-PSDB: AAD10325.
XX
PT Preventing or attenuating pathoangiogenic conditions e.g. cancer,
PT chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by
PT administering group B beta-hemolytic Streptococci toxin receptor or its
PT fragment
XX
PS Claim 4: Page 45-47; 52pp; English.
XX
XX The present sequence is full length group B beta-haemolytic Streptococci
XX (GBS) toxin receptor protein, HP59 from human. The present invention
XX relates to a method for preventing or attenuating a patho-angiogenic
XX condition in a mammal which comprises administering to the mammal one
XX or more GBS toxin receptors or their immunogenic fragments to induce
XX or maintain an immune response to one of GBS toxin receptors. The
XX method is useful for preventing or ameliorating pathoangiogenic
XX conditions such as cancer, scarring during wound healing, gliosis
XX during repair of nerve injury, chronic wounds, keloids, reperfusion
XX injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and
XX psoriasis in mammals. The proteins of the invention are also used
XX as vaccines.
XX
SQ Sequence 536 AA:

```

```

Query Match 100.0%; Score 2836; DB 22; Length 536;
Best Local Similarity 100.0%; Pred. No. 1.3e-283;
Matches 536; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MAAGATPPRPVOPARPGFGLSGRRSLCOVASTPAHVGVMSPYRDLARNDGEESTDR 60
DB 1 maagantpprpvoparpgfglsgrsllcqvastpahvgvmrspyrdlarndgeestdr 60
QY 61 TPLPGAPRAEAAPVCCSARYNLAIAFFGFTVVALRVNLVALVDMVDSNTTLEDNRT 120
DB 61 tplpgapraeaapvccsarynlaiaffgftvvalrvnlvalvdmvdsnttlednrt 120

```

```

DB 61 tplpgapraeaapvccsarynlaiaffgftvvalrvnlvalvdmvdsnttlednrt 120
QY 121 SRACREHSAPIKVNHNQTKKYYOWDAETOGMILGSPFYGIITQIPGVYASKIGKML 180
DB 121 sracrehsapikvnhnqtkkyyowdaetogmilkspfygiitqipgyvaskigkml 180
QY 181 GFGILCTAVLTLETPRIADLGVPGLVLRALLEGISGEVTFPPAMHAMSSWAPPLERSKLT 240
DB 181 gfgilctavltletpriadlgvpglvlrallengisgevtfppamhamsswapplersklt 240
QY 241 SISYACAGUQCTVSLPLSGITCYMMMTYVFRTGTIGFWELMLTWLGYSDTPQKHKRIS 300
DB 241 sisyacaguqctvslplsgitcymmmtvyfrrgtigfwelmltwlgyysdtpqkhkris 300
QY 301 HYKEYILSLRNQLSOKSVPMVPLIKSLPLAIVVAHPSYMTFTLTLPYMKEI 360
DB 301 hykeyilslrnqlsoksvpmvplikslplaivvahpsymtftltlpymkel 360
QY 361 LRFNVQENGFLSSLPYLGSWLWCMILSGQADNLRAMNFTLCVRRTFSLIGMIGPAVFL 420
DB 361 lrfnvqengflsslpylgswlwcmlsgqaadnlramnftlcvrtrfsligmigpavfl 420
QY 421 VAAFTGCDYSLAAVAFLTSTVLGFCSSGFSINHLDIAPSYAGILGTRNFPATIPGMV 480
DB 421 vaagftgcdyslaavafltstvlgfcssgfsinhldiapsyagilgtrnfatipgmw 480
QY 481 GPVIAKSLTPDNTVGEWQVTFYTAALINVGAFETFLFAKGEQONALDNHCHRH 536
DB 481 gpviaksltpdntvgewqvtfytaalinvgaftflfakgeqvonalndhghrh 536

RESULT 3
AA45087
ID AA45087 standard; Protein; 495 AA.
XX
AC AA45087;
XX
XX 31-MAY-2000 (first entry)
XX
DT Partial human GBS toxin receptor (HP55).
XX
DE Human GBS toxin receptor; group B beta-haemolytic streptococci; HP55;
KW pathological vascularisation; cancer metastases; angiogenesis;
KW neovascularisation; reperfusion injury; scarring; keloid;
KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW endothelial cell proliferation; antibacterial; anticancer.
KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH Region 14..20
FH FT /note= "Region of high hydrophobicity used to design
FH FT antigenic peptide"
FH FT /note= "Region of high hydrophobicity used to design
FH FT antigenic peptide"
FH FT /note= "Region of high hydrophobicity used to design
FH FT antigenic peptide"
FH Region 26..30
FH FT /note= "Region of high hydrophobicity used to design
FH FT antigenic peptide"
XX
PN WO200005375-A1.
XX
PD 03-FEB-2000.
XX
PF 22-JUL-1999; 99WO-US16676.
XX
PR 22-JUL-1998; 98US-0093843.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
PI Helieryist CG; Fu C;
XX

```

DR WPI: 2000-205377/18.
DR N-PSDB; AAZ50875.
XX
XX New polynucleotide encoding mammalian receptor for streptococcus toxin,
PT useful for diagnosis and treatment of, e.g. pneumonia in neonates -
XX
XX Claim 9; Page 22; 109pp; English.
XX
XX The present sequence is partial human GBS (group B beta-haemolytic
CC streptococci) toxin receptor (HP55). GBS toxin receptor is an integral
CC protein with seven transmembrane domains. Expression vectors comprising
CC the coding region can be transformed into host cells to express GBS
CC toxin receptor and its fragments. Detecting the receptor in tissues is
CC used to diagnose pathological vascularisation, e.g. for detecting cancer
CC metastases. GBS toxin receptors are useful for treating conditions
CC associated with pathological angiogenesis or neovascularisation
CC (specifically cancer, reperfusion injury, scarring during wound healing,
CC keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
CC neural injury), and to raise specific antibodies used for treating early
CC onset disease. Inhibitors of this receptor are useful for treating
CC pathological or hypoxia-induced endothelial cell proliferation and
CC migration.
XX
XX
SQ Sequence 495 AA:

Query Match 92.4%; Score 2621; DB 21; Length 495;
Best Local Similarity 100.0%; Pred. No. 1.7e-261;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 MRSVRLARDGEESTDRPLPGAPRAEAAPVCCSARNALALAFGEFFIYALRVNL 101
DB 1 mrsrvrlardgeestdrplpgapraeaapvccsarnalalafgffiyalrvnl 60
QY 102 SVALVDVDSNTLEDRKTSKACPEHSAPIKVHHNOTGKKYQMDAETOGWILGSFEFGYI 161
DB 61 svalvdmvdsnttledrtskacpehsapikvhhnqtgkkyqmdaetqgwilgsffgyi 120
QY 162 ITQIPGGYVASKIGKMLGFGILGTAVLTLPPIAADLGVPLIVLRALEGLGEGVTFP 221
DB 121 itqipggyvaskigkmlgfgilgtavltlfpiaadlvgpplivlraleglgegvtfp 180
QY 222 AMHAMSSMAPPLERSKLISYAGAOIGYISLPISGIICYNNMTYVYFFETGIGFW 281
DB 181 amhamsswapplersklisisyagagilgyvislpisgiicymnmtcyvffetgigifw 240
QY 282 FLIMIWLVSDTPQKHKRISHYEKEYILSLRNOLSSOKSVWPVILKSLPLMAIVAHFS 341
DB 241 flimwlvdsdtpqkhrishyekeyilslrnqlssqkswvpwplkslplwaihfs 300
QY 342 YNNWFTYLLTLPTMYKEILRFNVQENGFLSSLPLGSLWLCMLISGOAADNLRAKNFST 401
DB 301 ynnwftylltlptymkeilrfnvqengflsslplgswlcmllsggaadnlraknfst 360
QY 402 LCVRRIFSLIGMIGRAVFLVAAAGTIGCDYSLAAVFLTISTTLGFGSSGSINHLDIAPS 461
DB 361 lcvrrifsligmigrafvflvaagfigcdyslaavfltisttlgfgssgsinhlidiaps 420
QY 462 YAGILLGITWTFATIPGMVPAVLAKSLTPNTVGEOMOTVYIAAINVFGAIFETLPAKG 521
DB 421 yagillgitwtfatipgmwpaavlaksltpntvgeomotvyiaainvfgaifetlpaag 480
QY 522 EVQNALNDHHGHRH 536
DB 481 evqnalndhhghrh 495

RESULT 4
AAB66967
ID AAB66967 standard; protein; 495 AA.
XX
AC AAB66967;
XX

DT 18-APR-2001 (first entry)
XX
XX Human AST.
DE
XX Human; AST; nootropic; immunotropic; gene therapy; Salla disease;
KW anion and sugar transporter; anion-cation symporter;
KW Salla acid transporter.
XX
XX Homo sapiens.
OS
XX
XX EP1069184-A1.
PN
XX
XX 17-JAN-2001.
PD
XX
XX 16-JUL-1999; 99EP-0202341.
PF
XX
XX 16-JUL-1999; 99EP-0202341.
PR
XX
XX (ALKU) AKZO NOBEL NV.
PA
XX
XX WPI: 2001-193090/20.
DR N-PSDB; AAF55900.
XX
XX New human transporter gene implicated in Salla disease and lysosomal
PT Salla acid transport, useful in assays for identifying new drugs, or
PT diagnosing Salla acid transport defects related to mutations in the
PT transporter gene -
XX
XX Claim 1; Page 10-12; 20pp; English.
PS
XX
XX The present sequence is human Anion and Sugar Transporter (AST) protein
CC sequence. AST has significant homology with several members of the
CC anion-cation symporter (ACS) family of transporters. AST is implicated in
CC Salla disease, and is useful in screening assays for identifying new
CC drugs. Compounds identified via AST screening is useful for preparing a
CC pharmaceutical suitable as an activator or inhibitor of a Salla acid
CC transporter protein. The pharmaceutical may be used in Salla acid
CC associated diseases and CNS/immune related disorders.
XX
XX
SQ Sequence 495 AA:

Query Match 92.4%; Score 2621; DB 22; Length 495;
Best Local Similarity 100.0%; Pred. No. 1.7e-261;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 MRSVRLARDGEESTDRPLPGAPRAEAAPVCCSARNALALAFGEFFIYALRVNL 101
DB 1 mrsrvrlardgeestdrplpgapraeaapvccsarnalalafgffiyalrvnl 60
QY 102 SVALVDVDSNTLEDRKTSKACPEHSAPIKVHHNOTGKKYQMDAETOGWILGSFEFGYI 161
DB 61 svalvdmvdsnttledrtskacpehsapikvhhnqtgkkyqmdaetqgwilgsffgyi 120
QY 162 ITQIPGGYVASKIGKMLGFGILGTAVLTLPPIAADLGVPLIVLRALEGLGEGVTFP 221
DB 121 itqipggyvaskigkmlgfgilgtavltlfpiaadlvgpplivlraleglgegvtfp 180
QY 222 AMHAMSSMAPPLERSKLISYAGAOIGYISLPISGIICYNNMTYVYFFETGIGFW 281
DB 181 amhamsswapplersklisisyagagilgyvislpisgiicymnmtcyvffetgigifw 240
QY 282 FLIMIWLVSDTPQKHKRISHYEKEYILSLRNOLSSOKSVWPVILKSLPLMAIVAHFS 341
DB 241 flimwlvdsdtpqkhrishyekeyilslrnqlssqkswvpwplkslplwaihfs 300
QY 342 YNNWFTYLLTLPTMYKEILRFNVQENGFLSSLPLGSLWLCMLISGOAADNLRAKNFST 401
DB 301 ynnwftylltlptymkeilrfnvqengflsslplgswlcmllsggaadnlraknfst 360
QY 402 LCVRRIFSLIGMIGRAVFLVAAAGTIGCDYSLAAVFLTISTTLGFGSSGSINHLDIAPS 461
DB 361 lcvrrifsligmigrafvflvaagfigcdyslaavfltisttlgfgssgsinhlidiaps 420

QY 462 YAGILGINTNTPATIPGMVPIYAKSLTPDNTVGEQWTFVYIAAAINVGAIFFTLFANG 521
DB 421 yagillgintntfatipgmwgvpiaksltpdntvgewqtfvylaaainvgailfftlfakg 480
QY 522 EVQNNALNDHGHRRH 536
DB 481 evgnwalndhghrth 495

RESULT 5
AAM38959 standard; Protein; 495 AA.
AAM38959;
22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 2104.
Human; nocotropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyocrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic; chemokine; thrombolytic; drug screening; arthritis; inflammation; leukaemia.
Homo sapiens.
MO200153312-A1.
26-JUL-2001.
26-DEC-2000; 2000MO-US34263.
21-JAN-2000; 2000US-0488725.
25-APR-2000; 2000US-0552317.
09-JUL-2000; 2000US-0598042.
19-JUL-2000; 2000US-0620312.
03-AUG-2000; 2000US-0653450.
14-SEP-2000; 2000US-0662191.
19-OCT-2000; 2000US-0693036.
29-NOV-2000; 2000US-0727344.

(HYSE-) HYSEQ INC.
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D; Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J; Zhao QA, Zhou P, Goodrich R, Drmanac RT;
WPI: 2001-442253/47.
N-PSDB: AA158115.
Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -
Example 3; SEQ ID NO 2104; 10078pp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AAM38642-AAM42213) with nocotropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemia and C.N.S disorders.
Note: The sequence data for this patent did not form part of the printed

CC specification.
SQ Sequence 495 AA:
Query Match 92.2%; Score 2614; DB 22; Length 495;
Best Local Similarity 99.8%; Pred. No. 9, 2e-261;
Matches 494; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 42 MRSPVADLARNDEESTDTPILLPGAPRAEAPVCCSARYNLAIIIFGFTYARVNL 101
DB 1 mrsprvdlarndeestdtpillpgapraeapvccsarynlaiaaifgfivayalrvnl 60
QY 102 SYALVDMVDSNTTLEDPNRTSKACPEHSAPIKVHHNOTGKKYOMDAETQOMIIGSFYGYI 161
DB 61 syalvdmvdsnttleednrtskacpehsapikvhhnqtgkkywdaetqwgllgsffgyi 120
QY 162 ITQIPGGYVASKIGCKMLLGFGLGTAVALTLFTPLAADLGVPILVLRALLEGEGVTRP 221
DB 121 itqipggyvaskigckmllgfgllgtavaltftplaadlgvpllvlrallengvtrp 180
QY 222 AMHAMSSWAPPLERSKLLSISYAGQLGTVISLPLSGITCYMMNTYFYFFGTIGIFW 281
DB 181 amhamsswapplerskllsisyagqlgtvislplsgilcyymntvyffgtlgtifw 240
QY 282 FLUMIWLVSDFPQKRRKISHYEKEYILSSLRNQLSSQKSVPMVPIIKSLPLMAIVAHPS 341
DB 241 flumiwlvdfpqrkrrishyekeyilsslrnqlssqksgvpmvpilkslplmaivahps 300
QY 342 YNMFTFYTLTLPTYYKEILRFNVQENGLSLPYLGSLCMTLSGOAADNLRKWNFST 401
DB 301 ynmftfylltlptymkeilrfnvqengflslpylgswlcmllsgqaadnlrakwnfst 360
QY 402 LCVRRIFSLIGMIGRAVEFYAAGFTGCDYSLAVAFITITTTGGPCSSGSFINHDIAS 461
DB 361 lcvrrifsligmigravefyaaagftgdcyslaavafitittttggpcssgsfinhdiaps 420
QY 462 YAGILGINTNTPATIPGMVPIYAKSLTPDNTVGEQWTFVYIAAAINVGAIFFTLFANG 521
DB 421 yagillgintntfatipgmwgvpiaksltpdntvgewqtfvylaaainvgailfftlfakg 480
QY 522 EVQNNALNDHGHRRH 536
DB 481 evgnwalndhghrth 495

RESULT 6
AAV45088 standard; Protein; 495 AA.
AAV45088;
31-MAY-2000 (first entry)
Sheep GBS toxin receptor (SP55).
Sheep GBS toxin receptor (SP55).
Sheep GBS toxin receptor; group B beta-haemolytic streptococci; SP55; pathological vascularisation; cancer metastases; angiogenesis; sheep; neovascularisation; reperfusion injury; scarring; keloid; chronic inflammation; rheumatoid arthritis; psoriasis; neural injury; endothelial cell proliferation; antibacterial; anticancer; anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic.
Ovis sp.
Key Location/Qualifiers
226..252
FT Domain "Outer boundary of transmembrane domain"
232..248
FT Domain "note= \"Inner boundary of transmembrane domain\"
363..389
FT Domain "note= \"Outer boundary of transmembrane domain\"
369..385

| | | | | |
|----|---------------|---|----|--|
| FT | Domain | /note= "Inner boundary of transmembrane domain" | FT | /label= PKC_phospho_site |
| FT | Domain | /note= "Outer boundary of transmembrane domain" | FT | /note= "Predicted phosphorylation site" |
| FT | Domain | 458..474 | FT | 37..39 |
| FT | Domain | /note= "Inner boundary of transmembrane domain" | FT | /label= PKC_phospho_site |
| FT | Domain | 135..157 | FT | /note= "Predicted phosphorylation site" |
| FT | Domain | /note= "Outer boundary of transmembrane domain" | FT | 55..57 |
| FT | Domain | 137..153 | FT | /label= PKC_phospho_site |
| FT | Domain | /note= "Inner boundary of transmembrane domain" | FT | /note= "Predicted phosphorylation site" |
| FT | Domain | 42..58 | FT | 73..75 |
| FT | Domain | /note= "Predicted transmembrane domain" | FT | /label= PKC_phospho_site |
| FT | Domain | 328..345 | FT | /note= "Predicted phosphorylation site" |
| FT | Domain | /note= "Outer boundary of transmembrane domain" | FT | 97..99 |
| FT | Domain | 328..344 | FT | /label= PKC_phospho_site |
| FT | Domain | /note= "Inner boundary of transmembrane domain" | FT | /note= "Predicted phosphorylation site" |
| FT | Domain | 390..407 | FT | 254..256 |
| FT | Domain | /note= "Outer boundary of transmembrane domain" | FT | /label= PKC_phospho_site |
| FT | Domain | 390..406 | FT | /note= "Predicted phosphorylation site" |
| FT | Domain | /note= "Inner boundary of transmembrane domain" | FT | 269..271 |
| FT | Domain | 97..100 | FT | /label= PKC_phospho_site |
| FT | Domain | /note= "Putative amidation site" | FT | /note= "Predicted phosphorylation site" |
| FT | Modified-site | 59..62 | FT | /label= PKC_phospho_site |
| FT | Modified-site | /note= "Asn is N-glycosylated" | FT | /note= "Predicted phosphorylation site" |
| FT | Modified-site | 71..74 | FT | /label= PKC_phospho_site |
| FT | Modified-site | /note= "Asn is N-glycosylated" | FT | /note= "Predicted phosphorylation site" |
| FT | Modified-site | 77..80 | FT | 8..35 |
| FT | Modified-site | /note= "Asn is N-glycosylated" | FT | /label= p56a |
| FT | Modified-site | 95..98 | FT | /note= "Immunogenic peptide" |
| FT | Modified-site | /note= "Asn is N-glycosylated" | FT | 7..22 |
| FT | Modified-site | 225..228 | FT | /label= p55a |
| FT | Modified-site | /note= "Asn is N-glycosylated" | FT | /note= "Immunogenic peptide" |
| FT | Modified-site | 302..305 | FT | 71..84 |
| FT | Modified-site | /note= "Asn is N-glycosylated" | FT | /label= p57a |
| FT | Modified-site | 357..360 | FT | /note= "Immunogenic peptide" |
| FT | Modified-site | /note= "Asn is N-glycosylated" | PN | MO200005375-A1. |
| FT | Modified-site | 11..14 | XX | |
| FT | Modified-site | /label= CK2_phospho_site | PD | 03-FEB-2000. |
| FT | Modified-site | /note= "Predicted phosphorylation site" | XX | |
| FT | Modified-site | 73..76 | PF | 22-JUL-1999; 99WO-US16676. |
| FT | Modified-site | /label= CK2_phospho_site | XX | |
| FT | Modified-site | /note= "Predicted phosphorylation site" | PR | 22-JUL-1998; 98US-0093843. |
| FT | Modified-site | 79..82 | PA | (UYVA-) UNITV VANDERBILT. |
| FT | Modified-site | /label= CK2_phospho_site | PI | |
| FT | Modified-site | /note= "Predicted phosphorylation site" | XX | |
| FT | Modified-site | 259..262 | DR | WPI: 2000-205377/18. |
| FT | Modified-site | /label= CK2_phospho_site | DR | N-PSDB; AA250876. |
| FT | Modified-site | /note= "Predicted phosphorylation site" | XX | |
| FT | Modified-site | 452..455 | PS | Claim 10; Page 22; 109pp; English. |
| FT | Modified-site | /label= CK2_phospho_site | XX | |
| FT | Modified-site | /note= "Predicted phosphorylation site" | XX | |
| FT | Modified-site | 126..131 | CC | The present sequence is sheep GBS (group B beta-haemolytic |
| FT | Modified-site | /note= "Predicted myristylation site" | CC | streptococci) toxin receptor (SP55). Sheep GBS toxin receptor is an |
| FT | Modified-site | 162..167 | CC | integral protein with seven transmembrane domains. Expression vectors |
| FT | Modified-site | /note= "Predicted myristylation site" | CC | comprising the coding region can be transformed into host cells to |
| FT | Modified-site | 172..177 | CC | express GBS toxin receptor and its fragments. Detecting the receptor in |
| FT | Modified-site | /note= "Predicted myristylation site" | CC | tissues is used to diagnose pathological vascularisation, e.g. for |
| FT | Modified-site | 205..210 | CC | detecting cancer metastases. GBS toxin receptors are useful for treating |
| FT | Modified-site | /note= "Predicted myristylation site" | CC | conditions associated with pathological angiogenesis or |
| FT | Modified-site | 209..214 | CC | neovascularisation (specifically cancer, reperfusion injury, scarring |
| FT | Modified-site | /note= "Predicted myristylation site" | CC | during wound healing, keloids, chronic inflammation (rheumatoid |
| FT | Modified-site | 337..342 | CC | arthritis or psoriasis) or neural injury), and to raise specific |
| FT | Modified-site | /note= "Predicted myristylation site" | CC | antibodies used for treating early onset disease. Inhibitors of this |
| FT | Modified-site | 386..391 | CC | receptor are useful for treating pathological or hypoxia-induced |
| FT | Modified-site | /note= "Predicted myristylation site" | CC | endothelial cell proliferation and migration. |
| FT | Modified-site | 403..408 | XX | |
| FT | Modified-site | /note= "Predicted myristylation site" | SO | Sequence 495 AA: |
| FT | Modified-site | 423..428 | | |
| FT | Modified-site | /note= "Predicted myristylation site" | | |
| FT | Modified-site | 427..432 | | |
| FT | Modified-site | /note= "Predicted myristylation site" | | |
| FT | Modified-site | 17..19 | | |

Query Match 82.1%; Score 2329; DB 21; Length 495;
Best Local Similarity 86.3%; Pred. No. 2.4e-231;

| Matches | 427; | Conservative | 39; | Mismatches | 29; | Indels | 0; | Gaps | 0; |
|---------|------|---|-----|------------|-----|--------|----|------|----|
| QY | 42 | MRSPVBDLARNDGEESTDPTLLPGAPRAEAAPVCCSARYNLAIAAFGEFIYALRVNL | 101 | | | | | | |
| Db | 1 | mkspsvdlaepsdgsdgtplllqrapraepapvcasaylnlaiflfigfivyslrnl | 60 | | | | | | |
| QY | 102 | SVALVDMVDSNTTLEDNRTSKACPEHSAPIKVHHNOTGKKYQWDAETOGMILGSFFGYI | 161 | | | | | | |
| Db | 61 | svalvdmvdsnttakdnrtscyaehsapikvlnhgtgkkyrwdactgwlisgffgyi | 120 | | | | | | |
| QY | 162 | ITQIPGCIYAAKIGKMLGFGILGTAVALTFTRPAADGVEPLIVRALEGIGEVTP | 221 | | | | | | |
| Db | 121 | ltqipgyvaasrsgklllfiglftacatflftplraadfgvalvalralegigevtyp | 180 | | | | | | |
| QY | 222 | AMHAMSSNAPPLERSKLLISYAGQLGTVISLPSGIICYYMMNTVYFPGTGIFM | 281 | | | | | | |
| Db | 181 | amhamsswapplerskllisysagqlgtvavpilsyvalcyymnvclyfifigvllw | 240 | | | | | | |
| QY | 282 | FLIMWLVSDTPQKHKRISHYEKEYILSLRNQSSOKSVWVPILKSLPLMAIVVAHFS | 341 | | | | | | |
| Db | 241 | flimwlvdsdtpethktltpyekeyilslknglssqskvwpmlksplwlvahfs | 300 | | | | | | |
| QY | 342 | YNNFTFTLLTLPFTYKELIRFVQENGFLSLPYGSMICMLISQOADNLRANKNFST | 401 | | | | | | |
| Db | 301 | ynwftfylltltpymkeylrfnigngflsavpylgcwlcmllsqqaadnlrarnwfs | 360 | | | | | | |
| QY | 402 | LCVRRIFSLIGMTGPAVFLVAAFGICDYSIAVAFLTISTTGFCSSGFSINHLDIAPS | 461 | | | | | | |
| Db | 361 | lcvrrifsligmipalflvaagfigcdyslaafltlstcllgsctsgfslnhldlaps | 420 | | | | | | |
| QY | 462 | YAGILLGINTFATIPGMVGPVIAKSLTPDNTVGEMQTFYIAAIVNGAIFFTLFARK | 521 | | | | | | |
| Db | 421 | yagillgintfatipgmipalflvaagfigcdyslaafltlstcllgsctsgfslnhldlaps | 480 | | | | | | |
| QY | 522 | EYQNMALNDHGHHRH | 536 | | | | | | |
| Db | 481 | eyqnmalsdghghrn | 495 | | | | | | |

RESULT 7

AAE06519 standard; Protein: 495 AA.

AC AAE06519;

DT 16-OCT-2001 (first entry)

DE Sheep group B beta-haemolytic Streptococci toxin receptor (SP55) protein.

XX Sheep; group B beta-haemolytic Streptococci toxin receptor; GBS; SP55;

KM cytotoxic; vulnerability; antiatherosclerotic; osteopathic; vasotropic;

KM wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;

KM keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;

XX vaccine.

XX Ovis sp.

XX Key

FH Region

FT Region

FT Region

FT Region

FT Region

FT Region

FT Region

/note= "Fragment of intracellular domain of GBS toxin receptor"

MO200156598-A2.

09-AUG-2001.

02-FEB-2001; 2001WO-US03662.

02-FEB-2000; 2000US-0179870.

(UYVA-) UNIV VANDERBILT.

Hellerqvist CG;

WPI: 2001-488844/53.

N-PSDB: AAD10326.

Preventing or attenuating pathoangogenic conditions e.g. cancer, by chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by administering group B beta-hemolytic Streptococci toxin receptor or its fragment

Claim 8: Page 50-52; 52pp; English.

The present sequence is group B beta-haemolytic Streptococci (GBS) toxin receptor protein, SP55 from sheep. The present invention relates to a method for preventing or attenuating a patho-angogenic condition in a mammal which comprises administering to the mammal one or more GBS toxin receptors or their immunogenic fragments to induce or maintain an immune response to one of GBS toxin receptors. The method is useful for preventing or ameliorating pathoangogenic conditions such as cancer, scarring during wound healing, gliosis during repair of nerve injury, chronic wounds, keloids, reperfusion injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and psoriasis in mammals. The proteins of the invention are also used as vaccines.

Sequence 495 AA:

Query Match 82.1%; Score 2329; DB 22: Length 495;

Best Local Similarity 86.3%; Pred. No. 2.4e-231;

Matches 427; Conservative 39; Mismatches 29; Indels 0; Gaps 0;

| | | | | | | | | | |
|----|-----|---|-----|--|--|--|--|--|--|
| QY | 42 | MRSPVBDLARNDGEESTDPTLLPGAPRAEAAPVCCSARYNLAIAAFGEFIYALRVNL | 101 | | | | | | |
| Db | 1 | mkspsvdlaepsdgsdgtplllqrapraepapvcasaylnlaiflfigfivyslrnl | 60 | | | | | | |
| QY | 102 | SVALVDMVDSNTTLEDNRTSKACPEHSAPIKVHHNOTGKKYQWDAETOGMILGSFFGYI | 161 | | | | | | |
| Db | 61 | svalvdmvdsnttakdnrtscyaehsapikvlnhgtgkkyrwdactgwlisgffgyi | 120 | | | | | | |
| QY | 162 | ITQIPGCIYAAKIGKMLGFGILGTAVALTFTRPAADGVEPLIVRALEGIGEVTP | 221 | | | | | | |
| Db | 121 | ltqipgyvaasrsgklllfiglftacatflftplraadfgvalvalralegigevtyp | 180 | | | | | | |
| QY | 222 | AMHAMSSNAPPLERSKLLISYAGQLGTVISLPSGIICYYMMNTVYFPGTGIFM | 281 | | | | | | |
| Db | 181 | amhamsswapplerskllisysagqlgtvavpilsyvalcyymnvclyfifigvllw | 240 | | | | | | |
| QY | 282 | FLIMWLVSDTPQKHKRISHYEKEYILSLRNQSSOKSVWVPILKSLPLMAIVVAHFS | 341 | | | | | | |
| Db | 241 | flimwlvdsdtpethktltpyekeyilslknglssqskvwpmlksplwlvahfs | 300 | | | | | | |
| QY | 342 | YNNFTFTLLTLPFTYKELIRFVQENGFLSLPYGSMICMLISQOADNLRANKNFST | 401 | | | | | | |
| Db | 301 | ynwftfylltltpymkeylrfnigngflsavpylgcwlcmllsqqaadnlrarnwfs | 360 | | | | | | |
| QY | 402 | LCVRRIFSLIGMTGPAVFLVAAFGICDYSIAVAFLTISTTGFCSSGFSINHLDIAPS | 461 | | | | | | |
| Db | 361 | lcvrrifsligmipalflvaagfigcdyslaafltlstcllgsctsgfslnhldlaps | 420 | | | | | | |
| QY | 462 | YAGILLGINTFATIPGMVGPVIAKSLTPDNTVGEMQTFYIAAIVNGAIFFTLFARK | 521 | | | | | | |

```

|||||
Db 421 yaqlllgltntfatipgmigpilarstlpentlgewgtvfclfaaainvfgaifftlkag 480
QY 522 EVQNMALNDHHGHRH 536
Db 481 evqmwalsdngghrn 495

RESULT 8
AAV45090
ID AAV45090 standard; Protein; 495 AA.
XX
AC AAV45090;
XX
DT 31-MAY-2000 (first entry)
XX
DE Human/Sheep consensus GBS toxin receptor.
XX
KW Human/Sheep consensus GBS toxin receptor; GBS;
KW group B beta-hemolytic streptococci; pathological vascularisation;
KW cancer metastases; angiogenesis; neovascularisation; reperfusion injury;
KW scarring; keloid; chronic inflammation; rheumatoid arthritis; psoriasis;
KW neural injury; endothelial cell proliferation; antibacterial;
KW anticancer; anti-angiogenic; anti-inflammatory; anti-arthritis;
XX
OS Homo sapiens.
OS Ovis sp.
XX
FH Key Location/Qualifiers
FT 1..495
FT /note= "All the X's in the sequence correspond to
FT non-consensus sites"
XX
PN MO200005375-A1.
XX
PD 03-FEB-2000.
XX
PE 22-JUL-1999; 99MO-US16676.
XX
PR 22-JUL-1998; 98US-0093843.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
PI Hellerqvist CG, Fu C;
XX
DR WPI: 2000-205377/18.
DR N-PDB: AAZ50880, AAZ50881.
XX
PT New polynucleotide encoding mammalian receptor for streptococcus toxin,
PT useful for diagnosis and treatment of, e.g. pneumonia in neonates -
XX
PS Disclosure: Page 98-100; 109pp; English.
XX
XX The present sequence is the human/sheep consensus GBS (group B beta
XX -haemolytic streptococci) toxin receptor. Expression vectors comprising
XX the coding region can be transformed into host cells to express GBS
XX toxin receptor and its fragments. Detecting the receptor in tissues is
XX used to diagnose pathological vascularisation, e.g. for detecting cancer
XX metastases. GBS toxin receptors are useful for treating conditions
XX associated with pathological angiogenesis or neovascularisation
XX (specifically cancer, reperfusion injury, scarring during wound healing,
XX keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
XX neural injury), and to raise specific antibodies used for treating early
XX onset disease. Inhibitors of this receptor are useful for treating
XX pathological or hypoxia-induced endothelial cell proliferation and
XX migration.
XX
SQ Sequence 495 AA;

```

Query Match 59.8%; Score 1698; DB 21; Length 495;
 Best Local Similarity 67.8%; Pred. No. 3..3e-166;

```

Matches 335; Conservative 0; Mismatches 159; Indels 0; Gaps 0;
QY 42 MRSPVRDLARNDEGSDRPPPLGAPRAEAPVCCSARVNLALTAFGEFYIALFVNL 101
Db 1 mxspvxdaxxxxgeexdxrxxxxxxxaxrkexapccsarvnxaxlxfgffxyxzlvn1 60
QY 102 SVALVDWVDSNTTLEDNRTSKACPEHSAPIKVHHNOTGKKYQWDAETQGIILGSFFGYI 161
Db 61 xvxxvxxmdxetxxdrrxxcxebsaplkvxxxqtxgkyxxdaetqgylxxfxgy1 120
QY 162 ITQIPGIVASKTGGKMLLFGILGTAVLFLFPPIAADLGVPILYVRALEGUGEYTFP 221
Db 121 ltxlpggyvasxxxgxxxlgx1xxxaxxllfpxaadxxgxxxlxalexl9egx1xp 180
QY 222 AMHAMSSMAPLERSKLSISYAGAOLGTVISLPGIICYYMNMWTVYFPFGTIGIFW 281
Db 181 amhamsswapplerslxxlxayagaxlgtvxsllp1sgxlcyymnwcyvltxxfxxxgxxw 240
QY 282 FLMIWLVSDTPQKHKRISHYEKEYILSLRNOLSSOKSVWVPIILKSLPMAIYVAHFS 341
Db 241 fxxwixlvsxtpxxxhkxxxxyekxxlls1xnqxsqkxvpxxxxkxllpwaxvaxfs 300
QY 342 YNWFTYTLFLLPYMKELLRFNVQENGFLSLPYLGSMICMLSGAALNLRKKNFT 401
Db 301 ynwftfylvxllp1pxmkxxlrfnvqengflsxxpylxxw1cm1lxgqaadnlraxwnfst 360
QY 402 LCVRFISLIGMIGPAVLVAAGFTGCDYSLAVAF1FTTLLGFCSSGSINHLDTAPS 461
Db 361 xxvxxxfsl1xm1gpxxxflxxxxxxgcdy1xyxflx1s1xlgfcssgs1mhl1xaps 420
QY 462 YAGLLGTFNTFATIPGMVGVIAKSLTPDNTVGEMQTVFYIAALNVFGEFTLEPAKG 521
Db 421 yaqlllgltntfatipgmigpilarstlpentlgewgtvfclfaaainvfgaifftlkag 480
QY 522 EVQNMALNDHHGHRH 535
Db 481 evqmwxxxxdhxghr 494

RESULT 9
AAG65238
ID AAG65238 standard; Protein; 314 AA.
XX
AC AAG65238;
XX
DT 20-NOV-2001 (first entry)
XX
DE Human sodium dependent phosphate co-transfer protein 35.
XX
KW Human; sodium dependent phosphate co-transfer protein 35;
KW hypophosphaturia; hypercalcaemia; hypophosphataemic rickets;
KW nephritis; gene therapy.
XX
OS Homo sapiens.
XX
PN CN1298882-A.
XX
PD 13-JUN-2001.
XX
PE 06-DEC-1999; 99CN-0124217.
XX
PR 06-DEC-1999; 99CN-0124217.
XX
PA (SHEN-) SHENGYUAN GENE DEV CO LTD SHANGHAI.
XX
PI Mao Y, Xie Y;
XX
DR WPI: 2001-503367/56.
DR N-PDB: AAH79234.
XX
XX Human Na-dependent phosphate cotransporter 35 and its coding sequence -
XX Claim 1; Page 21(Disclosure); 28pp; Chinese.

```

XX The present invention provides the protein and coding sequences of human
 CC sodium dependent phosphate co-transfer protein 35. The sequences can be
 CC used in the treatment of hypophosphaturia, hypercalcaemia,
 CC hypophosphataemic rickets and nephritis. The present sequence is the
 CC protein of the invention.

XX Sequence 314 AA:

Query Match 59.2%; Score 1679; DB 22; Length 314;
 Best Local Similarity 100.0%; Pred. No. 1,6e-164;
 Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 MHAMSSMAPPLERSKLTISYAGAGLTIVISPLISGIIICYNNMTVYFFETGIGFWE 282
 DB 1 mhamsmapplerskltisysagagltivisplisgiiicynnmtvyffetgigifw 60
 QY 283 LMIWLVSDTPQKRIKRIHYEKEYILSLRNOLSSQKSVWPVPIKSLPLMALIVAHFSY 342
 DB 61 lmiwlvsdtpqkkrishyekeyilslrnqlsqsksvwpvpilkslplmalivahfsy 120
 QY 343 NMTFTYLLTLPTMYKEIIRFNVOENGFTSLSPYIGSNLCMILSGAANLRAKMFSTL 402
 DB 121 nmtfylltlptmykeiirfnvengftslspylgswlcmilsgaadanlrakmfstl 180
 QY 403 CVRIRISLIGMIGPAVELVAAGFIGDYSIAVAFTISTTLGFCSSGSINHLDIAPSY 462
 DB 181 cvrrisligmigpavflvaagfigcdyslavafisttlgfcsggsinhldiapsy 240
 QY 463 AGILGINTFATIPGVGPVIAKSLTPDNTVGEWQTVFYIAAINVGAIFETLFAKE 522
 DB 241 agilgintfatipgmvgpviaksltpdntvgewqtfvfyiaainvgaifetlfake 300
 QY 523 YONWALNDHGRH 536
 DB 301 yonwalandhgrh 314

RESULT 10

AAM40745
 ID AAM40745 standard; Protein; 309 AA.

XX AAM40745;

XX 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 5676.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

XX 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0553317.

PR 09-JUL-2000; 2000US-0596042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

XX Tang YR, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

DR N-PSDB; AA159901.

PT Novel nucleic acids and polypeptides, useful for treating disorders

PS such as central nervous system injuries -

XX Example 2; SEQ ID NO 5676; 10078pp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S. disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 309 AA:

Query Match 56.9%; Score 1614; DB 22; Length 309;
 Best Local Similarity 100.0%; Pred. No. 7.8e-136;
 Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 PPPVOFAPRPGFGISGRSILCOVASTPAHVGMKSPVRLARNDEESTDRTPLPGA 67
 DB 6 pppvofaprpfgfgisgrsllcqvastprahvwmkspvrdlarndeestdrtplpga 65

QY 68 PRAEAPVCCSARYNLAIIAFGFFIVYALRVNLGYALVDMVDSNTLEDNRTSKACPEH 127
 DB 66 praeapvccsarynlaiaffgffivyalrvnlgyalvdmvdsntlednrtskacpeh 125

QY 128 SAPIKVHNQGTGKQWDAETQGMILGSFFYGYITQIPGCVYASKYIGKMLIGFSLGT 187
 DB 126 sapikvhnqgtgkkywdaetqgmilgsffgyitqipgyvaskyigkmligfslgt 185

QY 188 AVLTFTFPIAADLGVGRLLVLRALBGLGCVPRANHAMSSMAPPLERSKLTISYAGA 247
 DB 186 avltftfpiaadlgyvgrllvrlalbgldgcvpranhamssmapplerskltisysaga 245

QY 248 QLGTVISLPLSGIICYNNMTVYFFETGIGFWEFLMIWLVSDTPQKRIKRIHYEKEYI 307
 DB 246 qlgtvislplsqliicynnmtvyffetgigifwflmiwlvsdtpqkkrishyekeyi 305

QY 308 LSSL 311

DB 306 lssl 309

RESULT 11

AAM3914
 ID AAM3914 standard; Protein; 277 AA.

XX AAM3914;

XX 06-NOV-2001 (first entry)

DE Human polypeptide, SEQ ID NO: 4069.

KW Human; full length cDNA; cDNA synthesis; oligo-capping.
 XX Homo sapiens.
 OS
 XX EP1130094-A2.
 PN
 XX 05-SEP-2001.
 PD
 XX 07-JUL-2000; 2000EP-0114089.
 PF
 XX 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 XX (HELI-) HELIX RES INST.
 PA
 XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 PI
 XX WPI; 2001-524255/58.
 DR N-PSDB; AAK94876.
 DR
 XX 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 PS
 XX Claim 8; SEQ ID NO 4069; 1380bp + sequence listing; English.
 PS
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5' - and 3' -ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a polypeptide
 CC encoded by a full length human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 XX
 XX Sequence 277 AA:
 SQ

Query Match 51.1%; Score 1450; DB 22; Length 277;
 Best Local Similarity 98.9%; Pred. No. 5.6e-141;
 Matches 274; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 42 MRSPVRLARNDEESTDRPLPGAPRAEAPVCCSARVNLALAFGEFTVYALRVNL 101
 DB 1 mrsypvrlarndgeestdrplpgapraeapvcarsaynlalaffgftiyalrvnl 60
 QY 102 SVALVVDWDSNTLEDNRTSKACPEHSAPTRVHNHNOTGKKYOMDAETOGWTLGSEFGYI 161
 DB 61 svaivvdwdsntlednrtskacpehsaptrvhnhngcgykqywdaeqgwltgsefgyi 120
 QY 162 ITQIPGVYVASKIGKMLGFGILGTAVLTLPPIADLGVGPVIVRALEGLGEGYTFP 221
 DB 121 itqipgyvaskigkmlgfgilgtavltlftpraadlgygpilviraaleglsegvtfp 180
 QY 222 AMHAMSSWAPPLERSKLISIVAGAOGLVYISLPISGIIYVNNWTVYVFFSGTIGTFW 281
 DB 181 amhamsswapplersklisivagaglgvtyislpisgilyymwvtyvfyfsgtligtfw 240
 QY 282 FLTWIWLVSPTPOKHKRISHYKEYEILSLNOLSSQ 318
 DB 241 fltwiwlvsdtpqkhkrishyekeyilslnngagvq 277

RESULT 12
 ABB58701
 ID ABB58701 standard; Protein; 559 AA.
 AC ABB58701;

XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 2895.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 KW
 OS Drosophila melanogaster.
 XX
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001MO-US09231.
 PF
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PR
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PMD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL02804.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS
 XX Disclosure; SEQ ID NO 2895; 21pp + sequence listing; English.
 PS
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB857737-AB872072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 XX
 XX Sequence 559 AA:
 SQ

Query Match 35.1%; Score 996.5; DB 22; Length 559;
 Best Local Similarity 39.8%; Pred. No. 1e-93;
 Matches 208; Conservative 91; Mismatches 171; Indels 53; Gaps 10;

QY 51 RNDEESTDR-----TPILPGAPRAEAPVC-CSARVNLALAFGEFTVYALRVNL 102
 DB 31 rsddeadereafesgerpplirsgaaehngcgpkrthfnglgtfvyamrvnl 90
 QY 103 VALVVDWDSNTLEDNRTSKACPEHSAPTRVHNH-----OTGKKYOMDAETOG 150
 DB 91 valvvdwdsntlednrtskacpehsaptrvhn-----otgkkyomdaetog 143
 QY 151 WILGSEFTGYITITQIPGVYVASKIGKMLGFGILGTAVLTLPPIADLGVGPVIVRA 210
 DB 144 lvgsefgyititqipgyvaskigkmlgfgilgtavltlftpraadlgygpilvira 203
 QY 211 LEGIGEGVTPPAMHAMSSWAPPLERSKLISIVAGAOGLVYISLPISGIIYVNNWTVY 267
 DB 204 legigegvtpamhamsswapplersklisivagaglgvtyislpisgilyymwvtyv 263
 QY 268 TYVYFEFTGIGFWLLIMLWVSDTPQKHKRISHYKEYEILSLRNQ----- 314
 DB 264 psatyfegtlgflwflfiamwlvdydkpdpriresereylerisvqvrlnqdaae 323
 QY 315 -----LSSQKSVWVPLKSLPLMAIVVAHFSTNWFYTLTLPLPYMKELRPNV 365
 DB 315 -----LSSQKSVWVPLKSLPLMAIVVAHFSTNWFYTLTLPLPYMKELRPNV 365

Db 324 egqdevsirapreepripwslltvsplvwaillltqcgqwaftqltelptymnllhfd 383
 Oy 366 QENGLSLSPYGSWLCMTLSCGADNLRKAKNNFSTLCVRRIFSLIGMTGPAYFLVAGF 425
 Db 384 gsnallnaapyltstfwvgiaccsaladmmrlarrysllnsyklwtvasvpslgllgily 443
 Oy 426 IGCDSLAVAFLLTISTTGGFCSSGGSSTNHLDIAPSYAGILLGITNTFATIPGMWP-VI 484
 Db 444 vgcddwvwtftmIagvgsfggavaygnqnmhlaIspryagltmyltcnsaanlcgflapyl 503
 Oy 485 AKSLPTDNTVCEGQVFVYFAAAINVGAIFLFLFAKGEVONNA 527
 Db 504 gliinhretltqwhlvfwlaaglnlaagnltiyllfaaeqsws 546
 RESULT 13
 ID ABB60525
 XX ABB60525 standard; Protein; 502 AA.
 AC ABB60525;
 XX 26-MAR-2002 (first entry)
 DT 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 8367.
 DE Drosophila melanogaster polypeptide SEQ ID NO 8367.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 PD 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2001; 2000US-191637P.
 XX 23-MAR-2001; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 XX N-PSDB; ABL04628.
 DR New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 XX interactions -
 XX
 PS Disclosure: SEQ ID NO 8367; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SO Sequence 502 AA:

| | | | | |
|---------------------------|-------|--|-----------|------------|
| Query Match | 35.0% | Score 992 | DB 22 | Length 502 |
| Best Local Similarity | 42.4% | Pred. No. 2.5e+93 | | |
| Matches 190; Conservative | 93 | Mismatches 149 | Indels 16 | Gaps 3 |
| QY | 80 | RNLAALAFEGGFIYALRVNLISALVLDVMDNTLLDNRISKACDPHSAPITKVHHNQTG | 139 | |
| | : : | : | : | : |

```

Db      38 ryivvllffiffnvnslrnlsvaivamcentrvfd-----adgnvsyqg-- 83
Qy      140 KRYQMDAETQOVMILGSEFFGYIITQIDRGVYASKIGCKMLGFICTAVVLTFTPIAD 1599
      : : : : : | | | | | : | | : | | : : | | | | | : | |
Db      84 -dfpmdsqkql|lssffyylltqlfgygygktnignivfctgibgctalltltlpmas 142
Qy      200 LGVGEPLIVRALLEGEGEVTPRRAMHMMSSWPALEKSLSSYAGAOLGYVISP|LG 2399
      : : : : : | | | | | : | | : | | : | | | | | : | |
Db      143 hslmelfvrliegffegvtctprghavwaratpplersmaslaiaqnyagvtvampcsg 2020
Qy      260 IICYYMNMVYVPEFGTIGIFMFLIMVLVSDPRQKHKRISHYEKEYILSSLRNQSOK 3199
      : | | | | | : | | : | | : | | : | | : | |
Db      203 flackygwsvfyftglqvtlwtvltwlvfvcageldrctfskeccdy|lqhtdygyskhv 2622
Qy      320 SVVPWPLIKSLPLMAIVVAHFSYMWPTVYTLTLPLPYMKETILRPNOENGFSSLPYLG 3799
      : | | | | | : | | : | | : | | | | | : | | : | |
Db      263 khpwratftmpfyaimashiseenwgyflltqtprsfirdlntfdlctg|lsavpylam 3222
Qy      360 WLCMTLSQADNLRBAKKNFESTLCVRRFLSIGIGAVFLVAAAGFICGYSLAVAF|LT 4399
      : : : : : | | | | | : | | : | | : | | : | | : | |
Db      323 gllavsvy|adw|qvk|v|ctt|vrr|n|f|n|c|a|f|a|c|t|v|f|m|u|l|t|a|l|-|d|p|t|s|v|s|t|l 3811
Qy      440 STTLGGFCSSGFSIN|H|D|I|A|S|V|A|G|I|L|G|T|N|T|F|A|T|I|P|G|W|G|P|I|A|S|L|P|D|N|V|G|M|Q|T 4399
      : | | | | | : | | : | | : | | | | | : | | : | | : | |
Db      362 avgl|ga|fawsg|fay|h|d|ia|p|q|hasv|img|y|n|t|a|t|a|p|v|s|p|l|t|g|y|v|v|n|t|q|s|d|e|w|l 4411
Qy      500 VEYIAAINVFGAIFETLFAKGEONMA 527
      : | | | | | : | | : | | : | | : | | : | |
Db      442 if|f|s|a|g|v|y|c|v|y|w|f|y|c|s|g|d|q|e|w|a 469

```

RESULT 14
ID ABB65873 standard; Protein; 502 AA.
XX ABB65873;
AC ABB65873;
XX 26-MAR-2002 (first entry)
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 24411.
DE Drosophila melanogaster polypeptide SEQ ID NO 24411.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX Drosophila melanogaster.
OS Drosophila melanogaster.
XX WO200171042-A2.
PN WO200171042-A2.
XX 27-SEP-2001.
PD 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
PF 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PA (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI MPI; 2001-656860/75.
DR N-PSDB; ABL09976.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PI interactions -
PT interactions -
XX Disclosure; SEQ ID NO 24411; 21pp + Sequence Listing; English.

Search completed: July 15, 2002, 08:24:01
Job time: 106 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 09:39:25 ; Search time 4323.72 Seconds

(without alignments)
9146.309 Million cell updates/sec

Title: US-09-776-865-1

Perfect score: 2930
Sequence: 1 gttcgtcgaagccctccccc.....ctcaaaaaaaaaaaaaa 2930

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13376207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estbta:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description |
|------------|-------|-------------|-----------|----|-------------|
| 1 | 972 | 33.2 | 985 | 9 | AL550137 |
| 2 | 775.4 | 26.5 | 787 | 10 | BG198416 |
| 3 | 712.6 | 24.3 | 754 | 10 | BI907284 |
| 4 | 710 | 24.2 | 729 | 10 | BF971208 |
| 5 | 648.2 | 22.1 | 700 | 10 | BE869819 |
| 6 | 626.2 | 21.4 | 689 | 10 | BG400588 |
| 7 | 610.2 | 20.8 | 665 | 10 | BG290613 |
| 8 | 609.6 | 20.8 | 676 | 10 | BE867611 |
| 9 | 599.6 | 20.5 | 626 | 10 | BF676817 |
| 10 | 575.6 | 19.6 | 661 | 10 | BG400668 |
| 11 | 572.2 | 19.5 | 736 | 10 | BI860521 |
| 12 | 554.6 | 18.9 | 931 | 10 | BE878141 |
| 13 | 524 | 17.9 | 541 | 10 | BF312727 |
| 14 | 523.8 | 17.9 | 801 | 10 | BI661062 |
| 15 | 521 | 17.8 | 524 | 10 | BM272092 |
| 16 | 491.2 | 16.8 | 499 | 9 | AL597124 |
| 17 | 477.6 | 16.3 | 711 | 10 | BI697765 |

| | | | | | | | | |
|---|----|-------|------|------|----|----------|----------|------------|
| C | 18 | 462 | 15.8 | 462 | 9 | A1075876 | A1075876 | oz25c12.x |
| C | 19 | 445.2 | 15.2 | 904 | 10 | BE124137 | BE124137 | 601760838 |
| C | 20 | 438.2 | 15.0 | 895 | 10 | BE539146 | BE539146 | 602053009 |
| C | 21 | 432 | 14.7 | 432 | 9 | AA258513 | AA258513 | zr59d01.r |
| C | 22 | 423.8 | 14.5 | 427 | 10 | BM271768 | BM271768 | ig938d05.x |
| C | 23 | 420.2 | 14.3 | 682 | 9 | BB613552 | BB613552 | BB613552 |
| C | 24 | 413 | 14.1 | 579 | 10 | N21319 | N21319 | yx53h06.s1 |
| C | 25 | 406.8 | 13.9 | 411 | 9 | AA258357 | AA258357 | zr59d01.s |
| C | 26 | 402.8 | 13.7 | 633 | 10 | BI817031 | BI817031 | UMN07B04 |
| C | 27 | 394.8 | 13.5 | 412 | 9 | AM293672 | AM293672 | UI-H-B12 |
| C | 28 | 385.2 | 13.1 | 645 | 9 | BB537525 | BB537525 | BB537525 |
| C | 29 | 384.4 | 13.1 | 458 | 10 | H71642 | H71642 | ys10d10.r1 |
| C | 30 | 383.8 | 13.1 | 518 | 9 | AA833297 | AA833297 | ud05d09.r |
| C | 31 | 381.8 | 13.0 | 630 | 10 | BB610013 | BB610013 | BB610013 |
| C | 32 | 367.8 | 12.6 | 381 | 10 | BF772376 | BF772376 | IL5-IT002 |
| C | 33 | 367.2 | 12.5 | 629 | 9 | BB664731 | BB664731 | BB664731 |
| C | 34 | 367.2 | 12.5 | 629 | 9 | BB664731 | BB664731 | BB664731 |
| C | 35 | 353.4 | 12.1 | 893 | 10 | BG541099 | BG541099 | 602570265 |
| C | 36 | 349.4 | 11.9 | 351 | 9 | A1245423 | A1245423 | qk29d03.x |
| C | 37 | 349.4 | 11.9 | 349 | 10 | BF589677 | BF589677 | naa08b02 |
| C | 38 | 348 | 11.8 | 450 | 10 | H71643 | H71643 | ys10d10.s1 |
| C | 39 | 345.8 | 11.7 | 1100 | 10 | BG866603 | BG866603 | 602785377 |
| C | 40 | 342.4 | 11.7 | 329 | 9 | A1079154 | A1079154 | oy44d08.s |
| C | 41 | 318 | 10.9 | 429 | 10 | BF563945 | BF563945 | UI-R-C4-a |
| C | 42 | 316.4 | 10.8 | 367 | 10 | H63685 | H63685 | yr55e01.r1 |
| C | 43 | 307.8 | 10.3 | 301 | 10 | N31254 | N31254 | yx53h06.r1 |
| C | 44 | 301 | 10.5 | 301 | 10 | N31254 | N31254 | yx53h06.r1 |
| C | 45 | 288.6 | 9.8 | 721 | 10 | B060500 | B060500 | B060500 |

ALIGNMENTS

RESULT 1
AL550137
LOCUS AL550137 LTI_NFL006.PL2 985 bp mRNA linear EST 16-FEB-2001
DEFINITION AL550137 LTI_NFL006.PL2 Homo sapiens cDNA clone CS0D1040YK17 5
PRIMER, mRNA sequence.
ACCESSION AL550137
VERSION AL550137.1 GI:12886813
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 985)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Genoscope - Centre National de Sequencage
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 985
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1040YK17"
/clone_lib="LTI_NFL006.PL2"
/tissue_type="Placenta"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifestech.com URL :
http://fulllength.invitrogen.com"

FEATURES

source

BASE COUNT
ORIGIN

236 a 224 c 224 g 300 t 1 others

Query Match 33.2%; Score 972; DB 9; Length 985;
 Best Local Similarity 99.7%; Pred. No. 3.2e-187;
 Matches 983; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 327 tftcggcgccgscgtccctctctctgcagagtgccgagtaacctgtcgaatgagcgta 386
 |||||
 Db 1 tftcggcgccgscgtccctctctctgcagagtgccgagtaacctgtcgaatgagcgta 60

QY 387 tgaagtcctcgtctcgaacctgcccgcgaacatgcccgaagagacagcagccgcacg 446
 |||||
 Db 61 tgaagtcctcgtctcgaacctgcccgcgaacatgcccgaagagacagcagccgcacg 120

QY 447 ctctctaccggcgcccccagggccgaagccgctcgaatgctcgtcgtctgtctatca 506
 |||||
 Db 121 ctctctaccggcgcccccagggccgaagccgctcgaatgctcgtcgtctgtctatca 180

QY 507 acttagcaatttggccttttggcttttctctctctctctctctctctctctctctct 566
 |||||
 Db 181 acttagcaatttggccttttggcttttctctctctctctctctctctctctctctct 240

QY 567 gttcgtcgttagtgatgatatgtagatcaatacaacttagaagaataagaactcca 626
 |||||
 Db 241 gttcgtcgttagtgatgatatgtagatcaatacaacttagaagaataagaactcca 300

QY 627 aggcgtgtccagagcatctctcctccataaagttcatcaatacaacggtgaagaagt 686
 |||||
 Db 301 aggcgtgtccagagcatctctcctccataaagttcatcaatacaacggtgaagaagt 360

QY 687 accaattggagatgagaagaactccaagatgatatctcgttccctttttatgtctaca 746
 |||||
 Db 361 accaattggagatgagaagaactccaagatgatatctcgttccctttttatgtctaca 420

QY 747 tccacaagattctcgtgagagatgcttgcgcagcaaaatagaggagaaatgcgtcagat 806
 |||||
 Db 421 tccacaagattctcgtgagagatgcttgcgcagcaaaatagaggagaaatgcgtcagat 479

QY 807 ttggagatccttggcaactgctcctcaccctgttcaactcccatgtctgcaagattagag 866
 |||||
 Db 480 ttggagatccttggcaactgctcctcaccctgttcaactcccatgtctgcaagattagag 539

QY 867 ttggaccactcattgtacctcagaagcactagaagagactagaaggaggtgttaacttccag 926
 |||||
 Db 540 ttggaccactcattgtacctcagaagcactagaagagactagaaggaggtgttaacttccag 599

QY 927 ccatgcatgcatgctgctctcttgggtcccccctcttgaagaagaacactcttagca 986
 |||||
 Db 600 ccatgcatgcatgctgctctcttgggtcccccctcttgaagaagaacactcttagca 659

QY 987 ttctgcatgagagacagccttggagacagtaattctctctctctctctctctctgaata 1046
 |||||
 Db 660 ttctgcatgagagacagccttggagacagtaattctctctctctctctctctctgaata 719

QY 1047 gctacatatgaattggtactatgctctctctctctctctctctctctctctctgaata 1106
 |||||
 Db 720 gctacatatgaattggtactatgctctctctctctctctctctctctctctctgaata 779

QY 1107 tctcttcttggatcgttgaatgtagtgaacaccacacacacacacacacacacacacac 1166
 |||||
 Db 780 tctcttcttggatcgttgaatgtagtgaacaccacacacacacacacacacacacacac 839

QY 1167 atgaaagaagaatacattctctcatcaatgaagaatcagcttctctcagaagaagtcagtc 1226
 |||||
 Db 840 atgaaagaagaagaatacattctctcatcaatgaagaatcagcttctctcagaagaagtcagtc 899

QY 1227 cgtgggtaccattttaaataccctgcacacttgggcatcgttagtgcacacttctct 1286
 |||||
 Db 900 cgtgggtaccattttaaataccctgcacacttgggcatcgttagtgcacacttctct 959

QY 1287 acaactggaactttatctattatg 1312
 |||||
 Db 960 acaactggaactttatctattatg 985

RESULT 2
 BG198416 787 bp mRNA linear EST 21-APR-2001
 LOCUS R517674 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 DEFINITION BG198416
 ACCESSION BG198416
 VERSION BG198416.1 GI:13720103
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 787)
 Harrington,J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
 Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
 Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith,
 E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher,
 J., Danzig,J., and Ducar,M.
 Creation of genome-wide protein expression libraries using random
 activation of gene expression
 Nat. Biotechnol. 19 (5), 440-445 (2001)
 JOURNAL 21227151
 MEDLINE
 COMMENT Contact: Scott J. Cain
 Athersys, Inc.
 3201 Carnegie Ave, Cleveland, OH 44115, USA
 Tel: 216 431 9900
 Fax: 216 361 9596
 Email: scain@atersys.com
 High quality sequence stop: 549.
 Location/Qualifiers
 1. 787
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Athersys RAGE Library"
 /cell_line="HT1080"
 /note="See 'Creation of Genome-wide Protein Expression
 Libraries using Random Activation of Gene Expression',
 Nature Biotechnology, in press. Note that even though the
 cell type indicated is HT1080, since a random activation
 method was used, these sequence tags are not necessarily
 expressed in HT1080 under normal circumstances."
 BASE COUNT 214 a 166 c 154 g 253 t
 ORIGIN

Query Match. 26.5%; Score 775.4; DB 10; Length 787;
 Best Local Similarity 99.2%; Pred. No. 2.8e-147;
 Matches 779; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1645 gtagctgtatcctctcctgagcacacaatacattgccactatccagaaatgttg 1704
 |||||
 Db 3 gtagctgtatcctctcctgagcacacaatacattgccactatccagaaatgttg 62

QY 1705 gccgcgtcattcgttaaaagtcgagccctgataaacaactgttggagaatggcaacgctgt 1764
 |||||
 Db 63 gccgcgtcattcgttaaaagtcgagccctgataaacaactgttggagaatggcaacgctgt 122

QY 1765 ccatattgtgtgtcattaaatggttttggctccatttcttaacattcgcgaagg 1824
 |||||
 Db 123 ccatattgtgtgtcattaaatggttttggctccatttcttaacattcgcgaagg 182

QY 1825 tgaagtacaaaactggcctcctcaatgatacattgacagacagactgaagaaccaata 1884
 |||||
 Db 183 tgaagtacaaaactggcctcctcaatgatacattgacagacagactgaagaaccaata 242

QY 1885 aataacccgtcctcattaaatgatttattatcaatgataacccctcaagtcctctgt 1944
 |||||
 Db 243 aataacccgtcctcattaaatgatttattatcaatgataacccctcaagtcctctgt 302

QY 1945 attgtgaagcattatgcttttttaattgtaactgtatgaatttaagccat 2004
 |||||
 Db 303 attgtgaagcattatgcttttttaattgtaactgtatgaatttaagccat 362

QY 2005 aatcatgaataatcactagcttcgcgaataataaatgacgtgtttaaataatgaata 2064
|||||
Db 363 AATCATGAATAATCACTACTATTGCGAATAATAATGAAGTGTATATATGAATA 422
|||||
QY 2065 tatgtgaagtaggaacttacttagtggtacatactgctgtcgtatgctggacaatga 2124
|||||
Db 423 TATGTAAAGCTAGAGACTTCTACTTATAGTTTACATACCTGCTGATGTCGGCAACATGA 482
|||||
QY 2125 agtaggaacgttctgtgtgtttttagggccactactaaaggaagagctgtgaacagacc 2184
|||||
Db 483 AGTAGAGACGTTCTGTGTATTTTATAGGCCATACATAAAGGAATGAGCTGAAGACGAGCC 542
|||||
QY 2185 tccctgaacttctgttaactaactagatgaattcttcaggtactgataaacacctgt 2244
|||||
Db 543 TCTGTATACCTTGTCTTATTTAATCAAGTGAATTTCTCAGGTACTGTATAACACCTGT 602
|||||
QY 2245 tttgttgcacttccctcaataaaatgtcagctctctctgacactgaacctcaactt 2304
|||||
Db 603 TCTGTTCACCTTGTCTCAATAAAATGTCTGCTCTCTGACGCTTACACCTCAAACTTT 662
|||||
QY 2305 agcactctgtgagctgcacacacactctatatttgcctgcctgacactgagtgaggga 2364
|||||
Db 663 AGCATCTCTGTGTGAGCTGCATCTCATGTATATTTGCTGTGCAACTGAGACTGAGGGGA 722
|||||
QY 2365 gtgtgcccaggaagctgtgcccagaactccctcctgtgctcaggtcagagtgagtgcccagcgt 2424
|||||
Db 723 GTGTGCCAGGAGCGTCCAGACACCTCCTTCTGCTTCAAGGTGACAGTCCAGTGT 782
|||||
QY 2425 ttatc 2429
|||||
Db 783 TTATC 787
|||||

RESULT 3
LOCUS B1907284 754 bp mRNA linear EST 16-OCT-2001
DEFINITION 603063858P1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5212766 5',
mRNA sequence.
ACCESSION B1907284
VERSION B1907284.1 GI:16170094
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 754)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMU at:
<http://image.llnl.gov>
Plate: LLM11534 row: d column: 15
High quality sequence stop: 750.
Location/Qualifiers
1. 754
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5212766"
/clone_11b="NIH_MGC_118"
/tissue_type="Leukocyte"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6; Site1: NotI; Site2: EcoRV
(destroyed); RNA source leukocytes from anonymous pool of
non-activated adult donors. Library is oligo-dT primed
and directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range

BASE COUNT 171 a 182 c 194 g 207 t
ORIGIN
Query Match 24.3%; Score 712.6; DB 10; Length 754;
Best Local Similarity 99.1%; Pred. No. 1.6e-134;
Matches 748; Conservative 0; Mismatches 4; Indels 3; Gaps 3;
1.2-3.3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 027. Note:
this is a NIH-MGC Library.
QY 310 gcccgaggagcttcggagcttcggcgccctctctctcgcaggtgagctacacc 369
|||||
Db 2 GGCCTGGGGCTTCGGGCTTCGGGCGCGCTCCCTTCTCTCCAGAGTGGCAGTACACC 61
|||||
QY 370 tgcacagtaggcgtcactgaagctctccggttcagagactgcccggaaagatgagga 429
|||||
Db 62 TCTCATAGTAGGCGTATAGAGTCTCCGGTTCGAGACCTGCGCCGGAACGATGCGAGGA 121
|||||
QY 430 gtagcagagccgacagccctctctccggggggcccaaggagccgctccagtg 489
|||||
Db 122 GAGCAGGAGCCG-ACGCTCTTCTACCGGGCCGCCACGCGGCGAAGCGCTCCAGTGTG 180
|||||
QY 490 ctgctctgcgttaacaacttagcaatttggccttttggtttcttctcatgtgtatgc 549
|||||
Db 181 CTGCTCTGCTCTTACAACTTAGCAATTTTGGCCTTTTGGTTTCTTCTCATTTGTGATGC 240
|||||
QY 550 attacgttgaatctgagtggtgctgtagtgatattgtagatccaatacaacttga 609
|||||
Db 241 ATTACGTGTGAATCTGAGTGTGCGCTTACTGATATGTGATGATCAATCAACTTTAGA 300
|||||
QY 610 agataagaacttccaagcgctgtccagagacttctgtcccaataaaagttcatcata 669
|||||
Db 301 AATATATGAACCTTCCAGGCGGTGTCCAGAGCACTTCTGCCATTAAGTTTCAATATA 360
|||||
QY 670 tcaaacggtlaagaaagtacaaatgagatgcagaaactcaagagatgattctcgttct 729
|||||
Db 361 TCAACGCGTAGAAGTACCATGAGATGAGATCAAACTCAAGATGATTCGTGGTTCCT 420
|||||
QY 730 ttttaattgctacatcatcaacagattctgagagatatgttcccaagaaatagggg 789
|||||
Db 421 TTTTATGAGCTACATCATCATCACAGATTCCTGAGAGATATGTGTGCAGCAAAATAGGGGG 480
|||||
QY 790 gaaatgtcgtcagatattggattcttggagcgtctgtgccacacctgttaccocat 849
|||||
Db 481 GAAATGCTGTAGGCATTTGGATCTTGGGCACTGCTGTCTCACCTGTTCACTCCCAT 540
|||||
QY 850 tgcgcagatttagagatgtagccactcatgttactcagagactagaagagactagga 909
|||||
Db 541 TCTCTCAGATTTAGAGATTGGAGTGGACCACTCATTTGTACTCAAGACACTAAGAGACTAGAGA 600
|||||
QY 910 ggggttactattccaagcagatgcatgcatgtgtctctcttgggtccctcccttgaag 969
|||||
Db 601 GGGTGTATATTTCCAGGCATCATGATGCTGTCTTCTTGGGGTCCCTCTTGAAG 660
|||||
QY 970 aagcaactcttcaattctgtagtagagagagcaagcttggagaagtaa-tttctctc 1028
|||||
Db 661 AAGCAAACTTCTTAGCATTTTCATATGACAGACAGAGCTTGGACAGTATTTCTTTC 720
|||||
QY 1029 ctcttcttgaaataattgctactatataattg 1063
|||||
Db 721 CTCTTCTGTGATTA-TTCTCTACTATATGATTGG 754
|||||

RESULT 4
LOCUS BF971208 729 bp mRNA linear EST 22-JAN-2001
DEFINITION 602273220P1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4361140 5',
mRNA sequence.
ACCESSION BF971208
VERSION BF971208.1 GI:12338423
KEYWORDS EST.

```

SOURCE      human.
ORGANISM     Homo sapiens
REFERENCE    Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgarbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              CDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LNL at:
              http://image.llnl.gov
              Plate: LLM10003 row: h column: 05
              High quality sequence stop: 699.

FEATURES
  source
    1..729
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="NIH_MGC_84"
      /tissue_type="adrenal cortex carcinoma, cell line"
      /lab_host="DH10B (phage-resistant)"
      /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site:1:
      NotI; Site:2: SalI; Cloned unidirectionally; Oligo df:
      primed. Average insert size 1.229 kb. Library enriched for
      full-length clones and constructed by Life Technologies.
      Note: this is a NIH_MGC Library."

BASE COUNT   206 a      149 c      136 g      238 t
ORIGIN
Query Match 24.2%; Score 710; DB 10; Length 729;
Best Local Similarity 100.0%; Pred. No. 5,5e-134;
Matches 710; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1669 cacaataacttgcacattccagaagaatggttgcccgctatttgctaaagtgcac 1728
DB 1 CACAAATACATTTGCCATATCCAGGAATGTTGGCCGCTATTCCTAAAGTCTGAC 60

QY 1729 cccgtataaacactgttgagaatggaacacgctgtctatctgctgtctatatt 1788
DB 1 CCGGATTAACACGTGTGGAATGGAACCGGTCTATATGCTGCTGCTATTAATGT 120

QY 1789 tttgtgtgccatttcttctacattatcgcaaaagtgaagtaacaaactgggctctca 1848
DB 1 TTTTGTGCGCATTTCTTTACATATTCGCCAAAGGTGAAAGTCAAAAACGTGGCTCTCAA 180

QY 1849 tgcataccatggacacagaaactgaaagaaacataataatcctgcctctattatgta 1908
DB 1 TGAATCACCATGACACAGAACTGAAAGAACCAATTAATATCTGCTCTATTAAAGTA 240

QY 1909 ttttatttcatgttaacctcaaatgaccttctgtattgtgaagatctatgtctt 1968
DB 241 TTTTATTATTCATGTAAGTCAAGTCCCTTGATGTGTAAGCATTTCAATGCTTT 300

QY 1969 tttaattgactgtatagatttttaaggcctataatacatgaataataactagttgc 2028
DB 301 TTTTAATTTGACTTGTATGTATGATTTTAAAGCCTATATATCAATAATACATGATGCC 360

QY 2029 agataataaatagaactgtttaaataatgaataataatgaagctagagcttctactt 2088
DB 361 AGAATATAATAAATGAACCTGTTTAATTAATGAATTAATTAAGCTAGAGCTTCTACTTT 420

QY 2089 aggtcaataacctgctgtagtcggaacaatagaatgaagaactgtctgattttt 2148
DB 421 AGGTTCACATACCTGCTGCTAGTCGGGCAACATGAAGTGAAGACAGTTCGTGATTTTT 480

QY 2149 tagggccataactaaaggaatgagctgaaacagacctcctgataacttgccttaataa 2208

```

```

DB 481 TAGGCGCATCTAAAGGAGATGAGCTGAACACACCTCTGATACCTTTGCTTAATTAA 540
QY 2209 ctgatataataatctcaaggtacacgataaacaactgttgttcttctctataaaaa 2268
DB 541 CTGATGATTAATATCTCAGGACTGATTAACACCTGTTGTTGTTCTACTTCCATAAAAA 600
QY 2269 ttgtcagctctctctgacacttagacctcaacttagacatctctgtgagctgcaccc 2328
DB 601 TTGTCACCTCTCTCTGACACTTAGACTCAACTTATAGCATCTCTGTGGAGCTGCCATCC 660
QY 2329 actgtataattgcgctggaactggaactggaaggaggtgtgccagcag 2378
DB 661 ACTGTATAATTTGCGCCCTGCACTGAGGAGGAGGTGCCCCAGGACAG 710

RESULT 5
BE869819 700 bp mRNA linear EST 20-OCT-2000
LOCUS 60144651F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3850740 5',
DEFINITION mRNA sequence.
ACCESSION BE869819 GI:10318595
VERSION BE869819.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 700)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgarbs-r@mail.nih.gov
          Tissue Procurement: ATCC
          CDNA Library Preparation: Life Technologies, Inc.
          CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNL at:
          http://image.llnl.gov
          Plate: LLM9570 row: e column: 13
          High quality sequence stop: 697.

FEATURES
  source
    1..700
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="NIH_MGC_65"
      /tissue_type="adenocarcinoma"
      /lab_host="DH10B (phage-resistant)"
      /note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
      Site:2: SalI; Cloned unidirectionally. Primer: Oligo df.
      Average insert size 1.8 kb. Library constructed by Life
      Technologies."

BASE COUNT   185 a      141 c      130 g      244 t
ORIGIN
Query Match 22.1%; Score 648.2; DB 10; Length 700;
Best Local Similarity 99.1%; Pred. No. 2e-121;
Matches 683; Conservative 0; Mismatches 3; Indels 3; Gaps 3;

QY 1408 tatgatccgtctgctgcaagctgctacaaattgaaggcaaaatgaaattttcaactt 1467
DB 1 TATGATCCCTGCTGTGCAAGCTGCTGACAAATTTAAAGGCAAAATTTTCAACTTT 60

QY 1468 atgtgtcgaagaattttagccttataagaatgactggaactgcaglattcctgttagc 1527
DB 61 ATGTGTCCGAGAAATTTTAGCTTATAGGAATGATTTGACCTGCAGTATTCCTGTGTAC 120

QY 1528 tgcgtgctcaattgctgtgattatctcttggcggctgttcttaactataatcaaac 1587

```

Db 121 TCGTGCTTCATGCTGATTAATTCCTTTGGCCGCTGCTTCTTAATAATCAACAAC 180
QY 1588 actggagagccttct 1647
Db 181 ACTGGAGGCTTTGCTCTCTCTGATTTAGCATCAACCACTGATTTGGCTCTTCGTA 240
QY 1648 tgcctgtaaccctccgagcgcacacacacacacacacacacacacacacacacacacac 1707
Db 241 TCGTGTATCCCTCCGCGGATCAACAATATCATTTGCCACTATTCAGAGATGTTGGGCG 300
QY 1708 cgtcatctgtaaaagctgagccctgataacacgtcttgagaaatggcaaacctgttcta 1767
Db 301 CCGTATGCTAAAGCTGACCCCTGATTAACACTGTGTGAGATGGCAAAACCGTGTCTTA 360
QY 1768 tatctcct-gctgctcttaagtcttctctctctctctctctctctctctctctctct 1826
Db 361 TATTCGCTGCTCTATTAATGTTTGGGCGCATTTCTTAACCTATTTGGCAAAAGGTC 420
QY 1827 aagtaaaaactgggctcctcaatgatacaccatgagacacagacacacacacacacacaa 1886
Db 421 AAGTCAAAACCTGGGCTCTCATATGATCAACCATGACAGACAGACACTGAAGAACATATA 480
QY 1887 taatccctcct 1946
Db 481 TATTCCTGCTCTATTAATGATTAATTTATTTATTCATGTAACCTCAAGCTCTGTAT 540
QY 1947 tctgtaagact 2006
Db 541 TGTGTAGCATCTATGCTTTTATTTTAAATGTAATGTAATGA-TTTTAAAGCCATATA 599
QY 2007 tcaatgaatatacact-aggctccagaataataaataagacgtctgttaattgataata 2065
Db 600 TCATGAATATATCAGTCAGTGGCAGAAATATAAATGAACCTGTCTTACTATGATATAT 659
QY 2066 atgtaagctagact 2094
Db 660 AAGTAAGCTAGGACTTCTTAACCTTAGGTTTC 688

RESULT 6
LOCUS BG400588 689 bp mRNA EST 12-MAR-2001
DEFINITION 60246424P1 NIH_MGC_75 Homo sapiens CDNA clone IMAGE:4592171 5',
mRNA sequence.
ACCESSION BG400588
VERSION BG400588.1 GI:13294036
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 689)
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rstraub@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM1330 row: b column: 12
High quality sequence stop: 687.
Location/Qualifiers
1..689
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4592171"
/clone_id="NIH_MGC_75"
/lab_host="DH10B (T1 phage-resistant)"

FEATURES
source

/note="Organ: kidney; Vector: PDNR-LIB (Clontech); Site 1:
SfiI (ggccgctcgcc); Site 2: SfiI (ggcctatagcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCCATTAATGCGC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCGAGCGCGCCGACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.65
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: This is a NIH-MGC Library."

BASE COUNT 193 a 134 c 135 g 227 t
ORIGIN
Query Match 21.4%; Score 626.2; DB 10; Length 689;
Best Local Similarity 98.1%; Pred. No. 3.8e-117;
Matches 676; Conservative 0; Mismatches 8; Indels 5; Gaps 4;
QY 1688 attcagaatggttgcccgctcattgctaaagctgacccctgataacacgtgtga 1747
Db 1 ATTCCAGGATGTGTGGGCCCTCATGTCTAATAAGTCAACCCCGATTAACACTGTGGA 60
QY 1748 gaatggcaaacggtgtctctatatactgctgctctctctctctctctctctctct 1807
Db 61 GAATGGCAAAACCGTGTCTAATATCTGCTGCTATTAATGTTTGGGCCATTTCTTT 120
QY 1808 aacattctgcgaaggggaagtaacaaactggtctcaatgatacaccatgagacaga 1867
Db 121 ACATATTGCGCAAAAGGTGAATGACAAACCTGGCTCTCAATGATCACCATGACACAGA 180
QY 1868 cactgaagaaacaaataatactctgctcctctctctctctctctctctctctctct 1927
Db 181 CACTGAAGAACCAATTAATATCTGCTGCTATTAATGATTTTATTAATCAATGAAC 240
QY 1928 ctcaaatgctctctgctatgtaagcaattctctctctctctctctctctctctct 1987
Db 241 CTCAAAAGTCCCTTGTATTTGTAAGCATTTCTATGTTTAAATGTAATGTAAT 300
QY 1988 agattttaagcctataatacagaatacctggttcgcaagaataataaataagacgt 2047
Db 301 AATTTTAAAGGCTATATATCATGAATATACATGTAATGTAATGTAATGTAATGTAAT 360
QY 2048 tcttaatatataataatgtaagctagactgactctctctctctctctctctctctct 2106
Db 361 TGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 420
QY 2107 gctagtcgggcaacatgaagtagaacagttctgt-tgatttttagggcactactaaag 2165
Db 421 GCTAGTCGGGCAACATGAAGTAGGACAGTTCGTGTGATTTTAAAGGCAATGAAGG 480
QY 2166 gaatgactgaaacagactcctctgactctctctctctctctctctctctctctctct 2225
Db 481 GAATGAGCTGAACAGACCTCTGATACCTTTGCTTAATTAAGATGAATGAATTTCTCA 540
QY 2226 ggtactgataaacacactgtgtgttctact--tctcctaaaaatgtgcactcctct 2283
Db 541 GGTACTGATTAACACACTGTGTGTGTTGTTGCTCATTAATAATGTCAGAGTCTCTCT 600
QY 2284 gacacttagacactcaacttta-gaatctctgttgagctgacatccactgataatctg 2342
Db 601 GACACTTGAACCTCAACCTTATGAGCATCTCTGTGAGGCTGCATCCAGTATTAATTGCG 660
QY 2343 ccttgcaactgactgagggagagtgctgc 2371
Db 661 CCGTGAACCTGACTGAGGAGGAGTGTGCC 689

RESULT 7
LOCUS BG290613 965 bp mRNA EST 21-FEB-2001
DEFINITION 602388889P1 NIH_MGC_93 Homo sapiens CDNA clone IMAGE:4517662 5',
mRNA sequence.
ACCESSION BG290613

VERSION BG290613.1 GI:13047751
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM human.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM10411 row: a column: 23
High quality sequence stop: 634.

FEATURES
source
1..965
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4517662"
/clone_lib="NIH_MGC_93"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; Oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."

BASE COUNT 281 a 218 c 223 g 243 t
ORIGIN

Query Match 20.8%; Score 610.2; DB 10; Length 965;
Best Local Similarity 85.9%; Pred. No. 1e-113;
Matches 762; Conservative 0; Mismatches 108; Indels 17; Gaps 7;

QY 1789 ttctgtgcatcttcttcaactatcgccaaagtgaaagtaacaaactgggctctcaa 1848
|||||
DB 4 ttttggtgccattttcttcaactatcgccaaagtgaaagtaacaaactgggctctcaa 63
|||||

QY 1849 tggatccacggagcaaacgtaaggaacataataatctgcctctataatgta 1908
|||||
DB 64 tgaaccacatgagacacgacactgaaggaacataataatctgcctctataatgta 123
|||||

QY 1909 ttctattatcatgttaacctcaagtgctctgtatgtgtaagcatctatgtctt 1968
|||||
DB 124 tttttattttatcattgtaacctcaagtgctctgtatgtgtaagcatctatgtctt 183
|||||

QY 1969 ttctaatgtactgtatgattttaagggctataatcaatgaataatcaatgttgc 2028
|||||
DB 184 ttttaattgtactgtatgattttaagggctataatcaatgaataatcaatgttgc 243
|||||

QY 2029 agataataatgaactgttataatgaataatgaataatgaactgttataatga 2088
|||||
DB 244 agataataatgaactgttataatgaataatgaataatgaactgttataatga 303
|||||

QY 2089 aggttcaacatactgcctgtagtcgggacacataatgaatgaactgttctgtatctt 2148
|||||
DB 304 aggttcaacatactgcctgtagtcgggacacataatgaatgaactgttctgtatctt 363
|||||

QY 2149 taaggcacaactaaaggaaatgagctgaacacagactcctgataccttgcctaataa 2208
|||||
DB 364 taaggcacaactaaaggaaatgagctgaacacagactcctgataccttgcctaataa 423
|||||

QY 2209 ctgatatgatatctcgaagtaactgaataaactgttgcctctctccataaaaa 2268
|||||
DB 424 ctgatatgatatctcgaagtaactgaataaactgttgcctctctccataaaaa 483
|||||

QY 2269 ttgtcaactctctctgacacttagacctcaacttagatcatctctgtgga-gctgacat 2327
|||||
DB 484 ttgtcactctctctgacacttagacctcaacttagatcatctctgtgga-gctgacat 543
|||||

QY 2328 cactgtataatctgcctggaactggaactggaactggaactggaactggaactggaactgga 2387
|||||
DB 544 cactgtataatctgcctggaactggaactggaactggaactggaactggaactggaactgga 602
|||||

QY 2388 actccctccctgcttgaagtgtaagtgtaagtgtaagtgtaagtgtaagtgtaagtgtaag 2444
|||||
DB 603 actccctccctgcttgaagtgtaagtgtaagtgtaagtgtaagtgtaagtgtaagtgtaag 662
|||||

QY 2445 ---gcccaagcgaactgtcactcttgcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2497
|||||
DB 663 gcccaagcgaactgtcactcttgcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 722
|||||

QY 2498 aatgtgtagataaagccctgagtaggaagaagaagtagatagatcactgtatgtctgtat 2557
|||||
DB 723 aatgtgtagataaagccctgagtaggaagaagaagtagatagatcactgtatgtctgtat 782
|||||

QY 2558 acatccccaactcttccctccagcaagaagataatgtcgtgcatgtaactgga 2617
|||||
DB 783 aacactaccccaactcttccctccagcaagaagataatgtcgtgcatgtaactgga 839
|||||

QY 2618 agaaaatgcgaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 2664
|||||
DB 840 --agaaaatgcgaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 884
|||||

RESULT 8
BE867611 676 bp mRNA linear EST 20-OCT-2000
LOCUS BE867611
DEFINITION 601443127F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3847279 5',
mRNA sequence.
VERSION BE867611
XREF BE867611 GI:10316387
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 676)
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM9561 row: e column: 08
High quality sequence stop: 676.

FEATURES
source
1..676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3847279"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 145 a 171 c 184 g 176 t
ORIGIN

Query Match 20.8%; Score 609.6; DB 10; Length 676;
 Best Local Similarity 98.2%; Pred. No. 1.4e-113;
 Matches 638; Conservative 0; Mismatches 9; Indels 3; Gaps 2;

QY 310 gccccggggttcggggtgctggcgccgctccctctctcgcaggtggcagttacacc 369
 DB 27 ggcggggggttcggggtcggggcgccgctccctctctcgcaggtggcagttacacc 86
 QY 370 tgcacagcagcgtcatagatgctcgcggttcgagacccgcccgaacagatggcagga 429
 DB 87 tgcacagcagcgtcatagatgctcgcggttcgagacccgcccgaacagatggcagga 146
 QY 430 gagcacgagcgcagcgcctctctacccggcgcccccagcgagcgcagcgcctcagtg 489
 DB 147 gagcacgagcgcagcgcctctctacccggcgcccccagcgagcgcagcgcctcagtg 206
 QY 490 cgtgctgctgcttaaacactttagcactttagcactttagcactttagcactttagc 549
 DB 207 cgtgctgctgctgcttaaacactttagcactttagcactttagcactttagcactttagc 266
 QY 550 attacgtgtaattcgtgagtggttcggttagtgatagtgatagtgatagtgatagtgat 609
 DB 267 attacgtgtaattcgtgagtggttcggttagtgatagtgatagtgatagtgatagtgat 326
 QY 610 agataaagaacttccaaagcgctgctcagaagcattcgtccataaaagttcatcata 669
 DB 327 agataaagaacttccaaagcgctgctcagaagcattcgtccataaaagttcatcata 386
 QY 670 tcaaacgggttaagaagttacaaatgggagtgagcagaactcaagaatggatctcgttcc 729
 DB 387 tcaaacgggttaagaagttacaaatgggagtgagcagaactcaagaatggatctcgttcc 446
 QY 730 ttt-ttatggctacatcacacacagatctcgtgaagatattgttcgaacaaata - 99 786
 DB 447 tttgttatggctacatcacacacagatctcgtgaagatattgttcgaacaaata - 99 506
 QY 787 ggggaaatgctgctagatgattggagatccctggaactgctgctcctccttcaactc 846
 DB 507 ggggaaatgctgctagatgattggagatccctggaactgctgctcctccttcaactc 566
 QY 847 catctgctcagatattaggaattggagacactcatttactcagaagcactagaagactagg 906
 DB 567 catctgctcagatattaggaattggagacactcatttactcagaagcactagaagactagg 626
 QY 907 agaggggttacaatttcacagcatgcatgcatgcatgcatgcatgcatgcatgcatgcat 956
 DB 627 agaggggttacaatttcacagcatgcatgcatgcatgcatgcatgcatgcatgcatgcat 676

RESULT 9
 BF676817 626 bp mRNA linear EST 21-DEC-2000
 LOCUS 6020843801 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4248811 5',
 DEFINITION mRNA sequence.
 ACCESSION BF676817
 VERSION BF676817.1 GI:11950712
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 TITLE NIH-MGC http://mgc.nci.nih.gov/
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1995)
 CONTACT: Robert Strausberg, Ph.D.
 E-mail: cgaabs-remail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov
 Plate: LILCM1067 row: 0 column: 20
 High quality sequence stop: 598.
 Location/Qualifiers

FEATURES

source

1. 626
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4248811"
 /clone_id="NIH_MGC_83"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: prostate; Vector: pDNR-LIB (Clontech);
 Site-1: SfiI (ggccgctggc); Site-2: SfiI (ggccatagcc
); 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
 sequence: 5'-ATCTAGAGCGCGGCGCGCATG-drr(30)-BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."
 BASE COUNT 157 a 126 c 123 g 220 t
 ORIGIN

Query Match 20.5%; Score 599.6; DB 10; Length 626;
 Best Local Similarity 98.4%; Pred. No. 1.5e-111;
 Matches 616; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 1303 taacttatgacattatgacctactatataagaagagatcctaaggttcaatgttcaaga 1362
 DB 1 TACTTATATGACATATATGCTACTATATATGAGAGATCTAAGGTTCATATGTTCAAGA 60
 QY 1363 gaatgggttttatctctccttctccttatttagctcttggtatagatctctgctg 1422
 DB 61 GAATGGGTTTATATCTTCAATGCTTATATGAGCTTGTGATATGATATCTGCTCG 120
 QY 1423 tcaagctgctgcaaatlltaaggcacaatggaaatttcaactttagtgcagaaat 1482
 DB 121 TCAAGCTGCTGACAAATTTAAGGCAAAATGGAATTTTCAACTTATATGCTTGCAGAA 180
 QY 1483 tttagccttataagaatgattggaactgagatctcctgtagctgctgcttcaatg 1542
 DB 181 TTTTACCTTATATGAGAAATGATGAGACTCAGATATCTGTAAGTCTGCTTCAATGG 240
 QY 1543 ctgtgatactcttggcgctgcttctccttaactatacaaacactgggggcttttg 1602
 DB 241 CTGTGATATATCTTGGCCGCTTCTTCTTAATATCAACACACTGGAGGCTTTTG 300
 QY 1603 ctctctgatttagcatcaaccatctgatatctcctctgtagtgcgtatcctct 1662
 DB 301 CTCTTCTGATTTAGCATATACCATCTGATATGCTCTTGTATGCTGATCTCTCT 360
 QY 1663 gggcactacaataatatttgcacactatccaggaatggttggcccgatctgtaaaag 1722
 DB 361 GGGCATCACAAAATACATTGGCCACTATTCAGGAATGTTGGCCCGCTATGCTAAAG 420
 QY 1723 tctgaacctctgtaaacactgctggtggaatgagcaaacctggttctatatgtgcgtgctat 1782
 DB 421 TCTGACCCCTGATACACGCTTGGAGAAATGSCAAACCGTTCATATATGCTGCTAT 480
 QY 1783 taatgttttggatcatttcttctacactatctgcgaagtggaagtgaacaactgggc 1842
 DB 481 TATATGTTTGGTGCATTTTCTTACACTATTCGCCAAAGGTGAAGTCAAAACTGGCC 540
 QY 1843 tctcaatgataccatgagacagacactgaaagaaat-aaataactgcgtcctat 1901
 DB 541 TCTCAATGATACCATGAGACAGACAGACAGAACTAAATATATCTGCTAT 600
 QY 1902 taatgtattttatattatcaatgaac 1927
 DB 601 TTAATGTTATTTATATCATATGTAAC 626

| | |
|------------|---|
| RESULT | 10 |
| LOCUS | BG400668 |
| DEFINITION | BG400668 661 bp mRNA EST 12-MAR-2001 602466341F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4592170 5' , |
| ACCESSION | BG400668 |
| VERSION | BG400668 |
| KEYWORDS | BG400668.1 GI:13294116 |
| SOURCE | EST . |
| ORGANISM | human. |
| | Homo sapiens |
| | Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; |
| | Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. |
| REFERENCE | 1 (bases 1 to 661) |
| AUTHORS | NIH-MGC http://mgc.nci.nih.gov/ . |
| TITLE | National Institutes of Health, Mammalian Gene Collection (MGC) |
| JOURNAL | Unpublished (1999) |
| COMMENT | Contact: Robert Strausberg, Ph.D. |

Tissue Procurement: CLONTECH Laboratories, Inc.
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LICM130 row: b column: 11
 High quality sequence start: 7
 High quality sequence stop: 656.

```

source
1. .661
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4592170"
/clone_lib="NHL MGC_75"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgcgctcgagc); Site_2: SfiI (ggcgctatggc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGCGCCATTATGGC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGCCGAGCGCGCCGACATG-drr(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.66
kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NHL MGC Library."
188 a 127 c 122 g 223 t 1 others
BASE COUNT
ORIGIN

```

| Query Match | 19.6% | Score 575.6 | DB 10 | Length 661 |
|-----------------------|---|--------------------|----------|------------|
| Best Local Similarity | 97.6% | Pred. No. 1.1e-106 | | |
| Matches 647 | Conservative 0 | Mismatches 10 | Indels 6 | Gaps 6 |
| QY 1688 | atccagaagatggttggccgcctacatgctaagaatcgcaccccgatacaactgtga | 1747 | | |
| Db 1 | ATTCCAGGAATGTTGGSCCCGCTATTTCCTAAATCTCGACCCCTGATTAACCTTTGGA | 60 | | |
| QY 1748 | gaa-tggcaaacgltgtctatattgtcgtcgtctctaaatgttlttttgglygcattttct | 1806 | | |
| Db 61 | GAAGGCGCAACCGTGTTCTAFATGTGCGTGCTATTAATG-TTTTGGTGCCATTTTCTT | 119 | | |
| QY 1807 | tacacattgcgcacaagtgtaagtaacaaactggcgcttcacatgatcacatgagacag | 1866 | | |
| Db 120 | TACACTATTCCGCCAAAGGTGAAGTACAAAACTGGGCGCTCTCATATGATTCACATGAGACAG | 179 | | |
| QY 1867 | acactggaagagcaacaataataatctctgcctcctatataatgtatatttatatcatgtga | 1926 | | |
| Db 180 | ACACTGAAGGAGCAACATTAATTAATCTCCGCTAATTAATGATTTTATTATCATGTGA | 239 | | |
| QY 1927 | ccctaaagtgcctctgtatgttgtaagcatctctatgccttttttaatgtacctgtat | 1986 | | |
| Db 240 | CCTCAAAAGTGCCCTTCTGATATGTGTGTAACCATTTCTATGCTTTTATTATGTACTTGTAT | 299 | | |

| | | | |
|----|------|--|------|
| Oy | 1987 | tagtlttttaaggcctataatcatatgaataataactagttgccaagaataataaataaact | 2046 |
| | | | |
| Db | 300 | tagatttttaagccctatratatcatgaanaatpactagttgccagaataatpaaataaact | 359 |
| Oy | 2047 | gtgtttaatlatgaataalatgtaagctagaactctacttta-ggttcacatactgcc | 2105 |
| Db | 360 | ggtgttaattatgacatpaaatgtstaagctatgacacttctacttttagggttcacataactggcc | 419 |
| Oy | 2106 | tgctatgtcgggacacatagaagtagaacagttctgttgatltttttaaggccatactaa-ag | 2164 |
| Db | 420 | tgcctagtcggggacacatgaaatgacagacacttctgtgatttttttaaggccacactaagag | 479 |
| Oy | 2165 | ggaatgagctggaacagagacctctctatacttctgtcttaattaaactagatgat-aattt | 2223 |
| Db | 480 | ggaatgagctggaacagacacttctctatgacttctgttaatttaaacatgaaatgaaatttct | 539 |
| Oy | 2224 | cagtagctgatataaacacctgtgtctgttctacttctctcaataaactgtgcagctctctct | 2283 |
| Db | 540 | cagtagctgatataaacacactgtgtgtgttcacttctgtctcaataanaatgtgtcag-tctctct | 598 |
| Oy | 2284 | gaacacttgaacctcaaacactttagcatctctgttggagcttgcataccatgtatataatttgcg | 2343 |
| Db | 599 | gacacttgaacctcaaacctttttgacatctctgtggagactgcataccacactgtatataatttgcg | 658 |
| Oy | 2344 | ctg 2346 | |
| Db | 659 | ctg 661 | |

| | |
|------------|--|
| RESULT | 11 |
| Locus | B1860521 |
| Definition | B1860521 736 bp mRNA linear EST 10-OCT-2001 60338678/F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5395772 5' , mRNA sequence. |
| Accession | B1860521 |
| Version | B1860521.1 GI:16001268 |
| Keywords | EST . |
| Source | human. |
| Organism | Homo sapiens |
| Reference | Eumayr/Ota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 736) NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) |
| Authors | Contact: Robert Strausberg, Ph.D. |
| Title | |
| Journal | |
| Comment | |

```

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ULNL at:
http://image.llnl.gov
Plate: LRAM12008 row: m column: 21
High quality sequence stop: 713.
Location/Qualifiers
1..736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5395772"
/cclone_lib="NIH_MGC_87"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: breast; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."

```


Query Match 19.5%; Score 572.2; DB 10; Length 736;
Best Local Similarity 99.3%; Pred. No. 5.3e-106;
Matches 585; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 620 acttccaaagcgtgtccaaagcattctgtccataaagttcataatacaacggt 679
|||||
Db 1 ACTTCCAAAGCGTCCAAAGCAATTCGTCCTCCATTAAGATTCAATCAATCAAGCGGT 60

QY 680 aagaatgacaatggatgacaaactcaagaatgattctcgtctctttttatgac 739
|||||
Db 61 AAGAAATTCCAATGGATGAGAACTCAAGATGGATTCGTTCTTTTATAGC 120

QY 740 taatcatcacacagattctcgtgagatattgtccagacaaataagagggaatgctg 799
|||||
Db 121 TACATCATCACACAGATTCTGTGAGATATGTTGCCAGCAAAATAGGGGGAATGCTG 180

QY 800 ctagaattggagatccttggcaactgtctcctcaacctgttcaactccttgcagat 859
|||||
Db 181 CTAGGATTTGGGATCCTTGCGCACTGCTGTCACCCCTTTCACCTCCATTTGCTGCAGAT 240

QY 860 ttagagattggaccactcatgttactcagaagcaactagaagagatgagatgttaca 919
|||||
Db 241 TTAGGAGTTGGACCACTATTGTACTCAGACACTAGAAAGACTAGGAGAGGgtttaca 300

QY 920 ttccagcattgcatgcatgtgtctcttggctccctcccttgaagaagcaact 979
|||||
Db 301 TTTCCAGCATGATGCATGTGTGTCTTGTGGGCTCCCTCTTGAAGAAGCAAACTT 360

QY 980 cttagacattcgtatgacagagacagcttggagacagtaattctctctcttctga 1039
|||||
Db 361 CTTAGATTTCAATATCAGAGACAGCTTGGGACAGTAAATTTCTCTCTTCTGGA 420

QY 1040 ataattgtactatagattgagattgtcttcttacttttttgatcttttgaata 1099
|||||
Db 421 ATTAATTTGCTACTATATGAAATTTGGACTTATGCTTTACT-TTTTGTGACTATTTGAAATA 479

QY 1100 tttagttctcttctgtgacatcgtgtagttagtgacacacacaaacaaagagaatt 1159
|||||
Db 480 TTTTGGTTCTTTTGTGTGATCTGTGTAGTTAGTGACACACCAAAACCAAGAAATTT 539

QY 1160 tcccatatgaaaaggaatacatcttctcaatcaataagaatcagctt 1208
|||||
Db 540 TCCCATTAAGAAAAGAAATACATTTTCATCATTAAGAAATCAGGTAT 588

RESULT 12
BE878141 931 bp mRNA linear EST 20-OCT-2000
LOCUS 601487630F1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3890071 5',
DEFINITION mRNA sequence.
ACCESSION BE878141
VERSION BE878141.1 GI:10326917
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 931)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: DCTD/DMP/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed By: The I.M.A.G.E. Consortium (ULNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ULNL at:
http://image.llnl.gov
Plate: LHAM9672 row: 1 column: 08
High quality sequence stop: 594.
Location/Qualifiers

FEATURES

source 1..931
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3890071"
/clone_11b="NIH_MGC_69"
/tissue_type="large cell carcinoma, undifferentiated"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."

BASE COUNT 251 a 239 c 276 g 165 t

ORIGIN

Query Match 18.9%; Score 554.6; DB 10; Length 931;
Best Local Similarity 99.0%; Pred. No. 2e-102;
Matches 579; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 2340 tcgacctgcaactgagatgagagagatgtgcccagagcagctgcgaagcactccctctg 2399
|||||
Db 1 TCOCCTGGCAACTGCACTAGAGGAGAGTGTGCCAGGCACTGCCAAGCACTCCCTCCTG 60

QY 2400 gcttcagggtcagagtgcccaagcgtttatcagaagcagatcccaagcccaagcagtg 2459
|||||
Db 61 GCTTCAGGGTCAAGAGTGGCCAGGTTTATCAGAGGAGGAGCATCCAGCCAGGCGAGTGT 120

QY 2460 cgactctcgtgtgtgtccttctctctgaggggtatcaatgtgtatgaataagcctgag 2519
|||||
Db 121 CGACTCTTGCGGTGGTGGCTTTCTCTGAGGGGCTATCAATGTGTGATTAAGCCCTGAG 180

QY 2520 taggcaagagcagtgagatccactgtcgtatgacttgaataatccctaaacttcccttc 2579
|||||
Db 181 TAGGCAAGAGCACTGAGATCCACTGCTATGCTTGTATCATCTTAACCTTCCCTTCC 240

QY 2580 cagcacaagagatatctgctgcatgcaactgcaaaagaaataatgcaagcagcgccgag 2639
|||||
Db 241 CAGCACAGAGGAGATA-TGCTGGGATGCAACTGCAAAAGAAATAATGCAAGGCG- CGGG 298

QY 2640 caagtggtcattatgctgtatcccaagcaacttggggggctgagtggtggcaatcatgag 2699
|||||
Db 299 CACGGGTGCTCATGCTGTATATCCAGCACTTGGGGGGCTGAGGTGATCATGTAG 358

QY 2700 atcagaagttcagaccagcctggccagcatgtgaaacccaactctctaaataata 2759
|||||
Db 359 ATCAGAGATTCCAGACCACTGGCCAGCATGTGAAACCCATCTCTTAATAATAATA 418

QY 2760 aaaaattagctggcgtgtgtgacggcgctgttaatcccaagatcagaagagctgaggt 2819
|||||
Db 419 AAAAATTAGCTGGGCGTGTGTGACGGGCGCTGTAAATCCAGATTAACAGAGGCTGAGGT 478

QY 2820 agggagatcacttgaacccctggagagtggaagtgtgacgtgaacccaagatcaagcactgga 2879
|||||
Db 479 AGGAGAAATCACTTGAACCTGGAGAGATGAAGTTGCAAGTGAACCAAGATACGCCACTGCA 538

QY 2880 ctccaagcctggcgatggagcagagactccaactcaaaaaaataa 2924
|||||
Db 539 CTCGAGCTGGGCGATGGAGCGAGACTCCAATCAAAAAAATAA 583

RESULT 13
BF312727 541 bp mRNA linear EST 21-NOV-2000
LOCUS BF312727
DEFINITION 601898247F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4127794 5',
ACCESSION BF312727
VERSION BF312727.1 GI:11260579
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 541)

FEATURES

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: L1CM1019 row: m column: 11
High quality sequence stop: 537.
Location/Qualifiers

FEATURES
source

1. 541
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4127794"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pORF7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(g). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
Location/Qualifiers

BASE COUNT 153 a 132 c 151 g 105 t
ORIGIN

Query Match 17.9%; Score 524; DB 10; Length 541;
Best Local Similarity 99.8%; Pred. No. 3.4e-96;
Matches 533; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2396 cctgtgctcagggtcagagtgccagcgcttatcatagagcgacatccaaagccagagcca 2455
DB 2 CCTGGCTTCAAGGTCACAGATGCCCCAGGCTTATAGAGCGACATCCAGCCAGAGCCA 61
QY 2456 gtgtcagactctcggtcgtgtgcttccctcctcaggggcatcaatgtgtagataaagccc 2515
DB 62 GTGTGCACTCTTCGGCTGCGCTCTTCTCTGAGGGGCTATCAATGTAGATTAAGCCC 121
QY 2516 tgaatagaagaagcagtgatgactgctatgtcttgatataactctaacttccc 2575
DB 122 TGAATAGGCAAGACGACATGATCCTGATGCTTGTATACATCTCAAACTTCC 181
QY 2576 ttcccaagacagagaaatattgctgtgcatgcaacctgcaaaaagaataatgcgaagcgc 2635
DB 182 TTCCACACACAGAGATATTGGCTGGCATGCAACCTGCAAAAGAAAAATGCCAGAGGCG 241
QY 2636 cgggacagcgtgtcattatgctcctcctcagcactttgggggctgaggtggcgaatca 2695
DB 242 CGGGCAGCGTGGCTCATGCTGTATATCCAGACATTTGGGGGGCTGAGTGGCGAATCA 301
QY 2696 tgaagataagaattcgagacagcctgccaagatlgltgaaaccccatctctaataaa 2755
DB 302 TGAATAGAGAGTTGCGAGACCAAGCTTGCCAGCATGGTGAACCCCATCTCTATAAAA 361
QY 2756 tacaanaaatatagctggtgctgtgacggcgccctgtaatcccaataaccagaagagt 2815
DB 362 TACAAAAAATTAGCTGGGCTGTGTGAGCGGCGCTGTAAATCCAGATACAGAGAGCTG 421
QY 2816 aggtagaagaatacacttgaaccttgagaggttggaagttgcagtgaaaccaagaatcacgcac 2875
DB 422 AGGTAGAGAGATCACTTGAACCTGGAGGTGGAAGTTGACGTGAACCAAGATCAGCGCAC 481
QY 2876 tgcactcca-gcctgggcgatggagcgagactccaactcaaaaaaaaaaaaaa 2930
DB 482 TGCACTCCAGGCGCTGGCGATGGAGCAAGATCCAACTCAAAAAA 537

RESULT 14
BI661062 801 bp mRNA linear EST 12-SEP-2001
LOCUS 603304362F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5350046 5',
DEFINITION mRNA sequence.

ACCESSION BI661062
VERSION BI661062.1 GI:15575298
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 801)
NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
AUTHORS Contact: Robert Strausberg, Ph.D.
TITLE Email: cgabs-remail.nih.gov
COMMENT Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furch Ph.D.

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Plate: L1AM11889 row: 1 column: 15
High quality sequence stop: 799.
Location/Qualifiers

FEATURES
source

1. 801
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5350046"
/clone_lib="NIH_CGAP_Mam4"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SpO6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: oligo dT. Average insert 2.5 kb. Library constructed by Life Technologies, catalog # 12018-016. Investigators providing samples: Lothar Hennighausen/Priscilla Furch, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP Library."
Location/Qualifiers

BASE COUNT 173 a 200 c 220 g 208 t
ORIGIN

Query Match 17.9%; Score 523.8; DB 10; Length 801;
Best Local Similarity 81.9%; Pred. No. 3.6e-96;
Matches 641; Conservative 0; Mismatches 137; Indels 5; Gaps 3;

QY 380 ggcgtcatgagtcctccggttcagacctgcccgcgaagatgcgagagacagcagac 439
DB 22 GCGGTATAGAGGCCCTCGTCCGGGCTCCGGGGGAACGACGATGAGAGACTCGGAC 81
QY 440 cgaacgctctttaccggcgcccaagcggaagcgctccagtgtgtctctgct 499
DB 82 AGACACCCGCTCTGCGCGGGCGCCGCAACGAGCGGCTCCAGTGTCTCTGCT 141
QY 500 cgttaactaactagaatttggccttttggcttcttcaatgtgtatgcatcgtgtg 559
DB 142 CGGTACAACTTACGATTTTGGCGTTCTGTGTTTCTCTCTATGCTTACGGGTG 201
QY 560 aatctagtgctgcttagtggatagtagatccaataccaacttaagaataa 619
DB 202 AACCTGAGGTGTGCGTTAGTGAACATGGTATCAATCAACTGACTGATTAATAGA 261
QY 620 acttcaagcgctgtccagagacttctgtccataaagtatcataatcaaaaggt 679
DB 262 ACGTCTAAGAGAGTGTGGGAACATCTGCCCCCATTAAGTTCACACATCAGAGGT 321

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: July 15, 2002, 09:44:20 ; Search time 592.65 Seconds
(without alignments)
8488.247 Million cell updates/sec

Title: US-09-776-865-1
Perfect score: 2930
Sequence: 1 gtctgcgtcgaagccctccccc.....ctcaaaaaaaaaaaaaaa 2930

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N_Geneseq_032802.*
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
25: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|-------------|
| 1 | 2930 | 100.0 | 2930 | 21 | AAZ50879 |
| 2 | 2930 | 100.0 | 2930 | 22 | AAAD10325 |
| 3 | 2602 | 88.8 | 2602 | 21 | AAZ50875 |
| 4 | 2493.2 | 85.1 | 2712 | 22 | AAK94876 |
| 5 | 2359.4 | 80.5 | 2512 | 22 | AAAF5900 |
| 6 | 2018.6 | 68.9 | 2670 | 22 | AAH79234 |
| 7 | 1570.8 | 53.6 | 2844 | 21 | AAZ50876 |
| 8 | 1570.8 | 50.7 | 2844 | 22 | AAAD10326 |
| 9 | 1484.8 | 50.7 | 1488 | 22 | AAI58115 |

| | | | | | | |
|------|--------|------|--------|----|-----------|--------------------|
| 10 | 1413 | 48.2 | 1485 | 21 | AAZ50880 | Human/Sheep consen |
| C 11 | 1405.6 | 48.0 | 1975 | 22 | AAH99626 | Human protein enco |
| 12 | 1305 | 44.5 | 1485 | 21 | AAZ50881 | Human/Sheep consen |
| 13 | 912.2 | 31.1 | 929 | 22 | AAI59901 | Human polynucleoti |
| 14 | 827.8 | 28.3 | 853 | 22 | AAK93901 | Human cDNA clone r |
| 15 | 775.6 | 26.5 | 838 | 22 | AAK92364 | Human cDNA 5'-end |
| 16 | 541 | 18.5 | 1975 | 22 | AAH99626 | Human protein enco |
| C 17 | 511.2 | 17.4 | 535 | 22 | AAK93107 | Human cDNA 3'-end |
| 18 | 434.8 | 14.8 | 798 | 23 | AAZ56219 | DNA encoding novel |
| 19 | 434.8 | 14.8 | 1066 | 23 | AAZ77186 | DNA encoding novel |
| 20 | 375 | 12.8 | 375 | 22 | AAH52158 | Human ATP protein |
| C 21 | 327.2 | 11.2 | 349 | 22 | AAI13566 | Human breast cance |
| 22 | 292 | 10.0 | 300 | 20 | AAI214190 | Human gene express |
| C 23 | 261.4 | 8.9 | 264 | 22 | AAI22435 | Human breast cance |
| C 24 | 257 | 8.8 | 267 | 22 | AAI24574 | Human breast cance |
| 25 | 253.2 | 8.6 | 3407 | 22 | AAZ99852 | Genomic sequence # |
| 26 | 253.2 | 8.6 | 3407 | 22 | AAK90223 | Human digestive sy |
| C 27 | 251 | 8.6 | 433 | 22 | AAI24174 | Human breast cance |
| 28 | 250.6 | 8.6 | 1992 | 22 | AAK85653 | Human immune/haema |
| 29 | 250.6 | 8.6 | 1995 | 22 | AAK85652 | Human immune/haema |
| C 30 | 249.8 | 8.5 | 3422 | 22 | AAK52406 | Human polynucleoti |
| C 31 | 248.6 | 8.5 | 3608 | 22 | AAK67271 | Human immune/haema |
| 32 | 248.6 | 8.5 | 3608 | 22 | AAK74891 | Human immune/haema |
| C 33 | 248.6 | 8.5 | 3608 | 22 | AAK83192 | Human immune/haema |
| C 34 | 248.4 | 8.5 | 3178 | 22 | AAK80500 | Human immune/haema |
| C 35 | 246.8 | 8.4 | 3178 | 22 | AAK80499 | Human immune/haema |
| 36 | 246.4 | 8.4 | 325791 | 22 | AAZ53104 | Human Oestrogen re |
| 37 | 246.2 | 8.4 | 384 | 22 | AAK67754 | Human immune/haema |
| 38 | 246.2 | 8.4 | 393 | 22 | AAK67752 | Human immune/haema |
| 39 | 246.2 | 8.4 | 393 | 22 | AAK67753 | Human immune/haema |
| 40 | 245.6 | 8.4 | 19371 | 22 | AAI05706 | Human reproductive |
| 41 | 245.6 | 8.4 | 27154 | 22 | AAI05708 | Human digestive sy |
| C 42 | 245.6 | 8.4 | 31952 | 22 | AAK89370 | Human immune/haema |
| C 43 | 245.6 | 8.4 | 48203 | 22 | AAK70161 | Human immune/haema |
| C 44 | 245.6 | 8.4 | 48203 | 22 | AAK81663 | Human immune/haema |
| 45 | 245.6 | 8.4 | 48203 | 22 | AAK82628 | Human immune/haema |

ALIGNMENTS

| | |
|----------|---|
| RESULT 1 | |
| AAZ50879 | standard; cDNA; 2930 BP. |
| ID | AAZ50879 |
| XX | |
| AC | AAZ50879; |
| XX | |
| DT | 31-MAY-2000 (first entry) |
| XX | |
| DE | Full length human GBS toxin receptor (HP59) cDNA. |
| XX | |
| KW | Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59; |
| KW | pathological vasculatization; cancer metastases; angiogenesis; |
| KW | neovascularisation; reperfusion injury; scarring; keloid; |
| KW | chronic inflammation; rheumatoid arthritis; psoriasis; neural injury; |
| KW | endothelial cell proliferation; antibacterial; anticancer; |
| KW | anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| FH | Key |
| FT | Location/Qualifiers |
| FT | 263..1873 |
| FT | /*tag= a |
| FT | /product= "Human GBS toxin receptor protein" |
| XX | |
| XX | WO200005375-A1. |
| PN | |
| XX | |
| XX | 03-FEB-2000. |
| PD | |
| XX | |
| PF | 22-JUL-1999; 99WO-US16676. |
| XX | |
| PR | 22-JUL-1998; 98US-0093843. |
| XX | |

PA (UYVA-) UNIV VANDERBILT.
XX Hellervast CG, Fu C;
XX
DR WPI; 2000-205377/18.
DR P-PSDB; AA145089.
XX
PT New polynucleotide encoding mammalian receptor for streptococcus toxin,
PT useful for diagnosis and treatment of, e.g. pneumonia in neonates -
XX
PS Claim 3; Page 89-93; 109pp; English.
XX
XX The present cDNA sequence encodes full length human GBS (group B beta
CC -haemolytic streptococci) toxin receptor (HP39). This sequence was cloned
CC by using human embryo lung cDNA library as template. Expression vectors
CC comprising this cDNA can be transformed into host cells to express GBS
CC toxin receptor and its fragments. Detecting the receptor in tissues is
CC used to diagnose pathological vascularisation, e.g. for detecting cancer
CC metastases. GBS toxin receptors are useful for treating conditions
CC associated with pathological angiogenesis or neovascularisation
CC (specifically cancer, reperfusion injury, scarring during wound healing,
CC keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
CC neural injury), and to raise specific antibodies used for treating early
CC onset disease. Inhibitors of this receptor are useful for treating
CC pathological or hypoxia-induced endothelial cell proliferation and
CC migration.
XX
XX Sequence 2930 BP; 729 A; 698 C; 681 G; 822 T; 0 other;

Query Match 100.0%; Score 2930; DB 21; Length 2930;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2930; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttcgtcgaagccctcccttaattatgtgaattcaagttcccaatcccgccgca 60
Db 1 gttcgtcgaagccctcccttaattatgtgaattcaagttcccaatcccgccgca 60
QY 61 gcccccatctcgtcgtcggcgaaggtgccccgaactttaagaagggtgcagga 120
Db 61 gcccccatctcgtcgtcggcgaaggtgccccgaactttaagaagggtgcagga 120
QY 121 gcgggagaagtcgtcgcgaacacggctcccgagcatgtagccggcggaagcgg 180
Db 121 gcgggagaagtcgtcgcgaacacggctcccgagcatgtagccggcggaagcgg 180
QY 181 gctcacttgcgcgaatccctacagaaactccgaactccgcttcccttaaccacaa 240
Db 181 gctcacttgcgcgaatccctacagaaactccgaactccgcttcccttaaccacaa 240
QY 241 gccagaattgccacacctaagatgagcgcgaggcgatataccacggcccgccgtcca 300
Db 241 gccagaattgccacacctaagatgagcgcgaggcgatataccacggcccgccgtcca 300
QY 301 gccagctcggccggggggtcgcggtcgcggcgccggcgtccctctcgcaggtgagc 360
Db 301 gccagctcggccggggggtcgcggtcgcggcgccggcgtccctctcgcaggtgagc 360
QY 361 gaatataccctgcacgttaagcgtagaggttcgcgttcgagaaacttgcgggaacga 420
Db 361 gaatataccctgcacgttaagcgtagaggttcgcgttcgagaaacttgcgggaacga 420
QY 421 tggcgaaggaagcagacgcagcctcttaccggcgcccaaggcgccgaagcgcg 480
Db 421 tggcgaaggaagcagacgcagcctcttaccggcgcccaaggcgccgaagcgcg 480
QY 481 tccagtgctgctcgtcgttacaactagaatttggccttttttggttcttcat 540
Db 481 tccagtgctgctcgtcgttacaactagaatttggccttttttggttcttcat 540
QY 541 tggatgatatacgtgtgaactgagtggtcgttagtgatagatagatcaataac 600
Db 541 tggatgatatacgtgtgaactgagtggtcgttagtgatagatagatcaataac 600

QY 601 aacttaagaataagaactcccaagcggtgccagagcattcgtcccaataaagt 660
Db 601 aacttaagaataagaactcccaagcggtgccagagcattcgtcccaataaagt 660
QY 661 tcatcataatacaacggttaagaagtaaccaatggatgcgaactcaagaatgatct 720
Db 661 tcatcataatacaacggttaagaagtaaccaatggatgcgaactcaagaatgatct 720
QY 721 cgggtccttttttaagcacaatcacaacagatcccgaggaatgttgcggca 780
Db 721 cgggtccttttttaagcacaatcacaacagatcccgaggaatgttgcggca 780
QY 781 aataggggggaaaatcgtcgttaagatttggatcccttgcacgcgtccctccctgt 840
Db 781 aataggggggaaaatcgtcgttaagatttggatcccttgcacgcgtccctccctgt 840
QY 841 caatccattgctgcagatttaagaattggacaatcaatgtataccaagactagaag 900
Db 841 caatccattgctgcagatttaagaattggacaatcaatgtataccaagactagaag 900
QY 901 actagagaaggggttaacattcccaacatgcatgcatgtgtcttcttggtccccc 960
Db 901 actagagaaggggttaacattcccaacatgcatgcatgtgtcttcttggtccccc 960
QY 961 tcttgaagaagaacaactcttagcatttcgtagcagggcagcagcttggagcaat 1020
Db 961 tcttgaagaagaacaactcttagcatttcgtagcagggcagcagcttggagcaat 1020
QY 1021 ttctctcctctcttcggaataattgctactatataatgtgaatttgacttatgcttact 1080
Db 1021 ttctctcctctcttcggaataattgctactatataatgtgaatttgacttatgcttact 1080
QY 1081 ttgttgactatggaatatttgggttcttcttgtagatctgtttagttagtacacac 1140
Db 1081 ttgttgactatggaatatttgggttcttcttgtagatctgtttagttagtacacac 1140
QY 1141 acaaaaaacaagaagaattcccatatgaagaagaataatcttcatcataagaaga 1200
Db 1141 acaaaaaacaagaagaattcccatatgaagaagaataatcttcatcataagaaga 1200
QY 1201 tcagcttcttcacagaagtcagtcggtggtgtaaccattttaaaatccctgcacttg 1260
Db 1201 tcagcttcttcacagaagtcagtcggtggtgtaaccattttaaaatccctgcacttg 1260
QY 1261 ggctacgttagtgcaacttcttacaacttgaacttttaacttttagcattatt 1320
Db 1261 ggctacgttagtgcaacttcttacaacttgaacttttaacttttagcattatt 1320
QY 1321 gcttactatatagaagagatccctaaggttcaatgttcaagaagaatgttttaactc 1380
Db 1321 gcttactatatagaagagatccctaaggttcaatgttcaagaagaatgttttaactc 1380
QY 1381 attgccttatttaggctctgtgtatgtagtacctgctcgttcaagctgcgtacaatt 1440
Db 1381 attgccttatttaggctctgtgtatgtagtacctgctcgttcaagctgcgtacaatt 1440
QY 1441 aagggcaaaatggaattttcaactttagtgttcgaagaatttttagccttaagaat 1500
Db 1441 aagggcaaaatggaattttcaactttagtgttcgaagaatttttagccttaagaat 1500
QY 1501 gattggaccctgcagatatccctgtagctcgtgcttcatgtggtgatatctcttgc 1560
Db 1501 gattggaccctgcagatatccctgtagctcgtgcttcatgtggtgatatctcttgc 1560
QY 1561 cgttgccttccataataccaacaacacttggaggcttcttcctcgtatagcat 1620
Db 1561 cgttgccttccataataccaacaacacttggaggcttcttcctcgtatagcat 1620
QY 1621 caaccatctggatatgtccctctgtatgctgtatccctcctgggggatacaataacat 1680
Db 1621 caaccatctggatatgtccctctgtatgctgtatccctcctgggggatacaataacat 1680

| | | | |
|----|------|--|------|
| OY | 1661 | tgacattccaggaatggtgtggcccgctaatctgtcaaaatctcgacctgtataaac | 1740 |
| Db | 1661 | tgcacattccaggaatggtgtggcccgctaatctgtcaaaatctcgacctgtataaac | 1740 |
| OY | 1741 | tgttgggaatgagcaaaccggtgtctatatttgctgtgtatatagtttttgtggcat | 1800 |
| Db | 1741 | tgttgggaatgagcaaaccggtgtctatatttgctgtgtatatagtttttgtggcat | 1800 |
| OY | 1801 | ttctttacactatctgcgcgaagtgaaatgacaaactggtcctcaatgatacaatg | 1860 |
| Db | 1801 | ttctttacactatctgcgcgaagtgaaatgacaaactggtcctcaatgatacaatg | 1860 |
| OY | 1861 | acaaagacactggaagagacccaataataatcctgtcctcataatgattttattatc | 1920 |
| Db | 1861 | acaaagacactggaagagacccaataataatcctgtcctcataatgattttattatc | 1920 |
| OY | 1921 | atgcaacctcaaaatgacctctgtattgtgtgaagcatctaatgcttttttaattgac | 1980 |
| Db | 1921 | atgcaacctcaaaatgacctctgtattgtgtgaagcatctaatgcttttttaattgac | 1980 |
| OY | 1981 | ttgtattagatctttaaagacctataatcatgaatactagtgtgcagataataaaa | 2040 |
| Db | 1981 | ttgtattagatctttaaagacctataatcatgaatactagtgtgcagataataaaa | 2040 |
| OY | 2041 | tgaacctgtgtttaattatataatataatgtaagctaggaactctactattagttcacatc | 2100 |
| Db | 2041 | tgaacctgtgtttaattatataatataatgtaagctaggaactctactattagttcacatc | 2100 |
| OY | 2101 | ctgcctctatgctcgagcaataatgaagtgaagacagctctgtatctttttaaggccatc | 2160 |
| Db | 2101 | ctgcctctatgctcgagcaataatgaagtgaagacagctctgtatctttttaaggccatc | 2160 |
| OY | 2161 | aaagggaaatgagctgtaaaacagacctctctatatccttctgtcttaattaaactgatatat | 2220 |
| Db | 2161 | aaagggaaatgagctgtaaaacagacctctctatatccttctgtcttaattaaactgatatat | 2220 |
| OY | 2221 | tctcaagtgactgtaaaacacacctgtgtgttcaacttctcctaataaaatgtgcagcttc | 2280 |
| Db | 2221 | tctcaagtgactgtaaaacacacctgtgtgttcaacttctcctaataaaatgtgcagcttc | 2280 |
| OY | 2281 | tctgaaccttagacctaaactttaacatctctctgtgagctgacatctcaactgataatt | 2340 |
| Db | 2281 | tctgaaccttagacctaaactttaacatctctctgtgagctgacatctcaactgataatt | 2340 |
| OY | 2341 | cgacctggcaactggaactgtaaaagaggtgtgcccagagcgtgccaagcactccctccg | 2400 |
| Db | 2341 | cgacctggcaactggaactgtaaaagaggtgtgcccagagcgtgccaagcactccctccg | 2400 |
| OY | 2401 | cttcaagtgccagagtgcgcaagcggttatcatagagcgacatccaaagccagagccagtgc | 2460 |
| Db | 2401 | cttcaagtgccagagtgcgcaagcggttatcatagagcgacatccaaagccagagccagtgc | 2460 |
| OY | 2461 | gacctctcggtgtgtgcttctctctcgaaagggctatcaatgtgtatgaataaagccctgagt | 2520 |
| Db | 2461 | gacctctcggtgtgtgcttctctctcgaaagggctatcaatgtgtatgaataaagccctgagt | 2520 |
| OY | 2521 | aggaagaagagtgagatgccactgctcatgtctgttatcatctctaaacttccctccc | 2580 |
| Db | 2521 | aggaagaagagtgagatgccactgctcatgtctgttatcatctctaaacttccctccc | 2580 |
| OY | 2581 | agcaagaaggaatatgtgctgtgcatagcctgcaaacctgcgaanaaaatgcaagcggccggcc | 2640 |
| Db | 2581 | agcaagaaggaatatgtgctgtgcatagcctgcaaacctgcgaanaaaatgcaagcggccggcc | 2640 |
| OY | 2641 | acgggtgctcatgtcctgtataatcccaactttggggggctgtgaggtgcgaatacatgaga | 2700 |
| Db | 2641 | acgggtgctcatgtcctgtataatcccaactttggggggctgtgaggtgcgaatacatgaga | 2700 |
| OY | 2701 | tcaagagttcgaagacagccctgtgcagacatgtgtgaaaccccatctctactataaaatacaa | 2760 |
| Db | 2701 | tcaagagttcgaagacagccctgtgcagacatgtgtgaaaccccatctctactataaaatacaa | 2760 |
| OY | 2761 | aaattatgctgtggcggtgtgtacagcgccctgttaatcccgatactccaaagagcgctgagtgta | 2820 |

| | | | |
|----------|--|---|------|
| Db | 2761 | aaatlaactcgtgagctggtgacgagcgccctgaatccacagatactcaggagctgaggtta | 2820 |
| Qy | 2821 | ggagaatcaacttgaacacttgggaggttggaagtctgacgtgaaccaagaatcacgcctactgcac | 2880 |
| Db | 2821 | ggagaatcaacttgaacacttgggaggttggaagtctgacgtgaaccaagaatcacgcctactgcac | 2880 |
| Qy | 2881 | tcacgactctggcgagatggagcgagacatccaaactcaaaaaaaaaaaaaaaaaa | 2930 |
| Db | 2881 | tcacgactctggcgagatggagcgagacatccaaactcaaaaaaaaaaaaaaaaaa | 2930 |
| RESULT 2 | | | |
| ID | AADI0325 standard; DNA: 2930 BP. | | |
| XX | AADI0325; | | |
| XX | AADI0325; | | |
| DT | 16-OCT-2001 (first entry) | | |
| XX | Human group B beta-haemolytic Streptococci toxin receptor (HP59) DNA. | | |
| DE | Human group B beta-haemolytic Streptococci toxin receptor; GBS; HP59; | | |
| KW | cytotoxic; vulnerability; antiatherosclerotic; osteopathic; vasotrophic; | | |
| KW | prevention; attenuation; pathoangiogenic condition; cancer; scar; | | |
| KW | wound healing; gliosis; nerve injury; chronic wound; reperfusion injury; | | |
| KW | keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis | | |
| KW | vaccine; ds. | | |
| XX | Homo sapiens. | | |
| OS | Homo sapiens. | | |
| XX | Key | | |
| PH | Location/Qualifiers | | |
| FT | 263..1873 | | |
| FT | /tag= a | | |
| FT | /product= "Human GBS toxin receptor protein (HP59)" | | |
| XX | W0200156598-A2. | | |
| PN | 09-AUG-2001. | | |
| XX | 02-FEB-2001; 2001WO-US03662. | | |
| PF | 02-FEB-2001; 2000US-0179870. | | |
| PR | (UYVA-) UNIV VANDERBILT. | | |
| XX | Hellervist CG; | | |
| PI | WPI: 2001-468844/53. | | |
| DR | P-PSDB; AAE06518. | | |
| XX | Preventing or attenuating pathoangiogenic conditions e.g. cancer, | | |
| PT | chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by | | |
| PT | administering group B beta-hemolytic Streptococci toxin receptor or its | | |
| PR | fragment | | |
| XX | Disclosure; Page 41-44; 52pp; English. | | |
| PS | The present sequence is a DNA encoding full length human group B beta- | | |
| XX | haemolytic Streptococci (GBS) toxin receptor protein, HP59. The present | | |
| CC | invention relates to a method for preventing or attenuating a patho- | | |
| CC | angiogenic condition in a mammal which comprises administering to the | | |
| CC | mammal one or more GBS toxin receptors or their immunogenic fragments | | |
| CC | to induce or maintain an immune response to one of GBS toxin receptors. | | |
| CC | The method is useful for preventing or ameliorating pathoangiogenic | | |
| CC | conditions such as cancer, scarring during wound healing, gliosis | | |
| CC | during repair of nerve injury, chronic wounds, keloids, reperfusion | | |
| CC | injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and | | |
| CC | psoriasis in mammals. The proteins of the invention are also used | | |
| CC | as vaccines. | | |
| XX | Sequence 2930 BP; 729 A; 698 C; 681 G; 822 T; 0 other; | | |

Query Match 100.0%; Score 2930; DB 22; Length 2930;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2930; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttcgttcgaagccctcccttaattatgtgcaattcaagtcgccactgcccgcgcga 60
Db 1 gtctgttcgaagccctcccttaattatgtgcaattcaagtcgccactgcccgcgcga 60
QY 61 gccccactcaactctgtctgtggaaggtgtgcccctgactttaagaagggtgcgaga 120
Db 61 gccccactcaactctgtctgtggaaggtgtgcccctgactttaagaagggtgcgaga 120
QY 121 gcggagaaaggttcgtccgaaacacggctcccgccatgtcgaacacggcgagagcg 180
Db 121 gcggagaaaggttcgtccgaaacacggctcccgccatgtcgaacacggcgagagcg 180
QY 181 gctcaatttgcgcaactcctcagaaactccagaaactcgtctccctagtccaacca 240
Db 181 gctcaatttgcgcaactcctcagaaactccagaaactcgtctccctagtccaacca 240
QY 241 gccagagttgccacactaagatgagcgggcgagatgaacacggccgcgcgtcca 300
Db 241 gccagagttgccacactaagatgagcgggcgagatgaacacggccgcgcgtcca 300
QY 301 gccagctcggcccgggggtctcgagctgtcgggcgcgccctctctcgtccaggtgc 360
Db 301 gccagctcggcccgggggtctcgagctgtcgggcgcgccctctctcgtccaggtgc 360
QY 361 gactgaacacgtctcagctgaagctcgtcgtcgtccggttcgagaactgtccggaacga 420
Db 361 gactgaacacgtctcagctgaagctcgtcgtcgtccggttcgagaactgtccggaacga 420
QY 421 tggcgagagagcgacggacacgcgctctctacacggcgccacagggcgagagcgcg 480
Db 421 tggcgagagagcgacggacacgcgctctctacacggcgccacagggcgagagcgcg 480
QY 481 tccagtgctgctgtctgtctcgttacaactttagcaatttggccttttggcttccat 540
Db 481 tccagtgctgctgtctgtctcgttacaactttagcaatttggccttttggcttccat 540
QY 541 tgtgtatgcatatgctgtgaatcgtgaagtgttgcgttagtgagatagatgaattacaat 600
Db 541 tgtgtatgcatatgctgtgaatcgtgaagtgttgcgttagtgagatagatgaattacaat 600
QY 601 aacttgaagataatagaacttccaagcggtgtccagagatctgtctccataaagt 660
Db 601 aacttgaagataatagaacttccaagcggtgtccagagatctgtctccataaagt 660
QY 661 tcatcataatcaaacgggtgaagaatacgaatggatgtcgaagaactcaagaatgtattct 720
Db 661 tcatcataatcaaacgggtgaagaatacgaatggatgtcgaagaactcaagaatgtattct 720
QY 721 cgggtctttttttatggtacatcatcacacagatcccgaggagatgtttgcaagaa 780
Db 721 cgggtctttttttatggtacatcacacagatcccgaggagatgtttgcaagaa 780
QY 781 aataggggggaaaaatgtctgaagatttggagatcctgtgcactgtgttccacccgtt 840
Db 781 aataggggggaaaaatgtctgaagatttggagatcctgtgcactgtgttccacccgtt 840
QY 841 caactccatgtgcagatattgaagatgtgacacatcatgtactcagagacataagag 900
Db 841 caactccatgtgcagatattgaagatgtgacacatcatgtactcagagacataagag 900
QY 901 actagagagaggtgttcaattccacatccagcagcagatgtgtctcttgggtctcccc 960
Db 901 actagagagaggtgttcaattccacatccagcagcagatgtgtctcttgggtctcccc 960
QY 961 tcttgaagaagaacaactcttagcatcttctgtacgagagcaagcttggagacgaat 1020
Db 961 tcttgaagaagaacaactcttagcatcttctgtacgagagcaagcttggagacgaat 1020

QY 1021 ttctctctctctcttggaaataattgtctactatgatgaattggacttatgtcttact 1080
Db 1021 ttctctctctctcttggaaataattgtctactatgatgaattggacttatgtcttact 1080
QY 1081 ttgtgtactaatgtgaataatttggttcttcttgttgaactcgtgttagtgaacacc 1140
Db 1081 ttgtgtactaatgtgaataatttggttcttcttgttgaactcgtgttagtgaacacc 1140
QY 1141 acaaaaacacagaagaattccccaattgaagaagaataactcttccatcataagaaga 1200
Db 1141 acaaaaacacagaagaattccccaattgaagaagaataactcttccatcataagaaga 1200
QY 1201 tcaagcttcttccacagaagtcaagtccgaagtccaagtccaagaagaatgtttatcttc 1260
Db 1201 tcaagcttcttccacagaagtcaagtccgaagtccaagtccaagaagaatgtttatcttc 1260
QY 1261 ggctatcgtatgttgcacacttttcttaacaacttgacttttaacttatgtacatatt 1320
Db 1261 ggctatcgtatgttgcacacttttcttaacaacttgacttttaacttatgtacatatt 1320
QY 1321 gccctactatagaagaagatcccaagttccaagtccaagaagaatgtttatcttc 1380
Db 1321 gccctactatagaagaagatcccaagttccaagtccaagaagaatgtttatcttc 1380
QY 1381 atgtccttattagagctctgtgtatgtatgatccctgtcgtcgaagctgtcgaacatt 1440
Db 1381 atgtccttattagagctctgtgtatgtatgatccctgtcgtcgaagctgtcgaacatt 1440
QY 1441 aagggcaaaaatgtgaattttccaacttatgttctgagaagaatttttagccttatagaat 1500
Db 1441 aagggcaaaaatgtgaattttccaacttatgttctgagaagaatttttagccttatagaat 1500
QY 1501 gattggacccgcagatctcctgttagctgtctgtcttcaatgtgctgatatcttggc 1560
Db 1501 gattggacccgcagatctcctgttagctgtctgtcttcaatgtgctgatatcttggc 1560
QY 1561 cgtgtcttccactaactatcaacaacacgtggagagcttctgtctctgtatagcat 1620
Db 1561 cgtgtcttccactaactatcaacaacacgtggagagcttctgtctctgtatagcat 1620
QY 1621 caaccatctgatatgtctctctgtatgttgttatctctctgtggacataacatcat 1680
Db 1621 caaccatctgatatgtctctctgtatgttgttatctctctgtggacataacatcat 1680
QY 1681 tgcacactatccagaaatgttggccgcgtcatgtcgaagaagctgaacccgtgataaac 1740
Db 1681 tgcacactatccagaaatgttggccgcgtcatgtcgaagaagctgaacccgtgataaac 1740
QY 1741 tgttggagaatggcaaacgtgttctatattgtctgtcgtctatgaatgttttgggcac 1800
Db 1741 tgttggagaatggcaaacgtgttctatattgtctgtcgtctatgaatgttttgggcac 1800
QY 1801 ttctcttaacatgttgcgaagaagtgaagtaacaagaactgggctcctaagtacaatag 1860
Db 1801 ttctcttaacatgttgcgaagaagtgaagtaacaagaactgggctcctaagtacaatag 1860
QY 1861 acacagacacttgaagaagaacaataaataatccggtcctctatgaatgtatttcaattac 1920
Db 1861 acacagacacttgaagaagaacaataaataatccggtcctctatgaatgtatttcaattac 1920
QY 1921 atgtaaactcaaaatgtcctctgtatgttgaagcatctcattgtctttttaaattgac 1980
Db 1921 atgtaaactcaaaatgtcctctgtatgttgaagcatctcattgtctttttaaattgac 1980
QY 1981 ttgtattagaattttaaaggcttaataatgaataatcaactagttgcgaagaataaaga 2040
Db 1981 ttgtattagaattttaaaggcttaataatgaataatcaactagttgcgaagaataaaga 2040
QY 2041 tgaactgttcttaataatgaataatgttaagcttaagacttcaacttagttcacatat 2100
Db 2041 tgaactgttcttaataatgaataatgttaagcttaagacttcaacttagttcacatat 2100
QY 2101 ctgctgtctagttcgggcaacatgaagtaaggaacgttctgttgaatttttaaggacat 2160


```

DB 2101 ctgcccctgagctggcgaacatgaagtagacagctctgtatcttttagggccatct 2160
QY 2161 aaagggaagagctggaacagacccctcctgataccttctgcttaataatagatgaat 2220
DB 2161 aaagggaagagctggaacagacccctcctgataccttctgcttaataatagatgaat 2220
QY 2221 tctcagglactataaacaccgtgtgttctcacttctcctcaataaatgtcagctcc 2280
DB 2221 tctcagglactataaacaccgtgtgttctcacttctcctcaataaatgtcagctcc 2280
QY 2281 tctgcaacttagaacctcaaaactttagcatctctgtgtgagctgcacatcgtataatt 2340
DB 2281 tctgcaacttagaacctcaaaactttagcatctctgtgtgagctgcacatcgtataatt 2340
QY 2341 gcgcctggcgaactggagctggaggtgtgtccagagagctgccaagactccctccctgg 2400
DB 2341 gcgcctggcgaactggagctggaggtgtgtccagagagctgccaagactccctccctgg 2400
QY 2401 ctccagagctcagagtgccagcgtttatcagagcagcatccaagccagagccagtgctc 2460
DB 2401 ctccagagctcagagtgccagcgtttatcagagcagcatccaagccagagccagtgctc 2460
QY 2461 gactcttcggtgtgtgtccttccctctgaggggtatcaatgtgtagataaagccctgagt 2520
DB 2461 gactcttcggtgtgtgtccttccctctgaggggtatcaatgtgtagataaagccctgagt 2520
QY 2521 agggcgaagcagtgagatccactgtctgtttagataactcaaatctccctccc 2580
DB 2521 agggcgaagcagtgagatccactgtctgtttagataactcaaatctccctccc 2580
QY 2581 agcacagagaaatattgtgtgtcgtacgtcaaccctgcgaataaagaaatgtcgaagccgggc 2640
DB 2581 agcacagagaaatattgtgtgtcgtacgtcaaccctgcgaataaagaaatgtcgaagccgggc 2640
QY 2641 acggtgtgtcgtcgtgtgtatccaccagacttggggggtgtgtgtgtgtgtgtgtgtgtgt 2700
DB 2641 acggtgtgtcgtcgtgtgtatccaccagacttggggggtgtgtgtgtgtgtgtgtgtgtgt 2700
QY 2701 ttagaggttcgagagcagcctgtgccaagcagatgtgtgaaccctctctactaaataataga 2760
DB 2701 ttagaggttcgagagcagcctgtgccaagcagatgtgtgaaccctctctactaaataataga 2760
QY 2761 aaatagcttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2820
DB 2761 aaatagcttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2820
QY 2821 ggaagaatcacttgaaccttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2880
DB 2821 ggaagaatcacttgaaccttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2880
QY 2881 tccagccttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2930
DB 2881 tccagccttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2930

```

RESULT 3

AAZ50875
ID AAZ50875 standard; cDNA; 2602 BP.

AAZ50875;

31-MAY-2000 (first entry)

Partial human GBS toxin receptor (HP55) cDNA.

Human GBS toxin receptor; group B beta-haemolytic streptococci; HP55;
pathological vasculitis; cancer metastases; angioneurosis;
neovascularisation; reperfusion injury; scarring; keloid;
chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
endothelial cell proliferation; antibacterial; anticancer;
anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriasis; ss.

```

OS Homo sapiens.
XX Key Location/Qualifiers
XX Key 58..1545
XX CDS /tag=a
XX FT /product="Partial human GBS toxin receptor"
XX PN WO200005375-A1.
XX PD 03-FEB-2000.
XX PF 22-JUL-1999; 99WO-0516676.
XX PR 22-JUL-1998; 98US-0093843.
XX PA (UUYA-) UNIV VANDERBILT.
XX PI Helleqvist CG, Fu C;
XX DR WPI, 2000-205377/18.
XX DR P-PSDB; AAY45087.
XX PT New polynucleotide encoding mammalian receptor for streptococcus toxin,
XX PS useful for diagnosis and treatment of, e.g. pneumonia in neonates -
XX Claim 3; Page 77-80; 109pp; English.
XX
CC The present cDNA sequence encodes partial human GBS (group B beta
CC -haemolytic streptococci) toxin receptor (HP55). This sequence was cloned
CC by using human embryo lung cDNA library as template. Expression vectors
CC comprising this cDNA can be transformed into host cells to express GBS
CC toxin receptor and its fragments. Detecting the receptor in tissues is
CC used to diagnose pathological vascularisation, e.g. for detecting cancer
CC metastases. GBS toxin receptors are useful for treating conditions
CC associated with pathological angiogenesis or neovascularisation
CC (specifically cancer, reperfusion injury, scarring during wound healing,
CC keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
CC neural injury), and to raise specific antibodies used for treating early
CC onset disease. Inhibitors of this receptor are useful for treating
CC pathological or hypoxia-induced endothelial cell proliferation and
CC migration.
XX
SQ Sequence 2602 BP; 672 A; 576 C; 583 G; 771 T; 0 other;

Query Match 88.8%; Score 2602; DB 21; Length 2602;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2602; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 tcggcgccgagcctccctctctgcccaggtggcgaatcacctgtcagctagcgctcatg 388
DB 1 tcggcgccgagcctccctctctgcccaggtggcgaatcacctgtcagctagcgctcatg 60
QY 389 aggtctccggttcgagacctggcccgaaagatggtcgaagagagcagcagccagcct 448
DB 61 aggtctccggttcgagacctggcccgaaagatggtcgaagagagcagcagccagcct 120
QY 449 ctctaccggcgccccaagggcggaagccgctccaggtgtgtgtgtgtgtgtgtgtgtgtgt 508
DB 121 ctctaccggcgccccaagggcggaagccgctccaggtgtgtgtgtgtgtgtgtgtgtgtgt 180
QY 509 ttagaacttggccttttttggttctctcatgtgtgatacatgatacgttgtaactgagt 568
DB 181 ttagaacttggccttttttggttctctcatgtgtgatacatgatacgttgtaactgagt 240
QY 569 gttgcgttagtgatagtgtagatccaataacacttgaagaataatgaactccaag 628
DB 241 gttgcgttagtgatagtgtagatccaataacacttgaagaataatgaactccaag 300
QY 629 ggtgtgccgaagatctgtctccataaagttcaatcaataacaggtgaagaatgac 688
DB 301 ggtgtgccgaagatctgtctccataaagttcaatcaataacaggtgaagaatgac 360

```

| | | | |
|----|------|--|------|
| QY | 689 | caatgagatgagaagaactcaagatagatcttggtcttctttttttatgctcaactac | 748 |
| Db | 361 | caatgagatgagaagaactcaagatagatcttggtcttctttttttatgctcaactac | 420 |
| QY | 749 | acacagatctcttggaaggataltgttcgcaagaataatgggggaaatgtgtcaagatt | 808 |
| Db | 421 | acacagatctcttggaaggataltgttcgcaagaataatgggggaaatgtgtcaagatt | 480 |
| QY | 809 | gggattccttggaactgctgtctcctcaaccctgtttcaatccattgctgcagattaaagt | 868 |
| Db | 481 | gggattccttggaactgctgtctcctcaaccctgtttcaatccattgctgcagattaaagt | 540 |
| QY | 869 | ggagcaactcttgcaccagacagatcgtagaagaagcctgagaggggtgttattctccacc | 928 |
| Db | 541 | ggagcaactcttgcaccagacagatcgtagaagaagcctgagaggggtgttattctccacc | 600 |
| QY | 929 | atgcattgcacgtgtctctcttggtctccccctcttgaaagagcaactcttagact | 988 |
| Db | 601 | atgcattgcacgtgtctctcttggtctccccctcttgaaagagcaactcttagact | 660 |
| QY | 989 | tcgtatgcagagacagacgcttggagacglaattctctctctcttcttctgaaataattgc | 1048 |
| Db | 661 | tcgtatgcagagacagacgcttggagacglaattctctctctcttcttctgaaataattgc | 720 |
| QY | 1049 | tactatagaatttgacttatgtctctactatttttggfacatcgtgaatatttgtgtt | 1108 |
| Db | 721 | tactatagaatttgacttatgtctctactatttttggfacatcgtgaatatttgtgtt | 780 |
| QY | 1109 | ctttgtgtagctctgtagttagttagacacacacaaacacaaagaaattccattat | 1168 |
| Db | 781 | ctttgtgtagctctgtagttagttagacacacacaaacacaaagaaattccattat | 840 |
| QY | 1169 | gaaaaggaatacatcttctcaatcaatgaagaatcagcttcttcaacagaagtcagtcgg | 1228 |
| Db | 841 | gaaaaggaatacatcttctcaatcaatgaagaatcagcttcttcaacagaagtcagtcgg | 900 |
| QY | 1229 | tgggtaaccattttaaaaatccctgcgaacttgggcatagtagtgtgcacactttcttac | 1288 |
| Db | 901 | tgggtaaccattttaaaaatccctgcgaacttgggcatagtagtgtgcacactttcttac | 960 |
| QY | 1289 | aactggaacttttataacttatatgacataatbctccactataatagaagatccctaag | 1348 |
| Db | 961 | aactggaacttttataacttatatgacataatbctccactataatagaagatccctaag | 1020 |
| QY | 1349 | ttcaatgttcaagagaatbvggtttatccttcaatgcttatttgaagcctbgtgtatgt | 1408 |
| Db | 1021 | ttcaatgttcaagagaatbvggtttatccttcaatgcttatttgaagcctbgtgtatgt | 1080 |
| QY | 1409 | atgataccgtctgttcaagctgctgcgaacaatttaagggcaaaatgtgaattttcaactta | 1468 |
| Db | 1081 | atgataccgtctgttcaagctgctgcgaacaatttaagggcaaaatgtgaattttcaactta | 1140 |
| QY | 1469 | tgtttctgcgaaattttttagcctctatagaagaatgtggccgcgaagatctccgttagct | 1528 |
| Db | 1141 | tgtttctgcgaaattttttagcctctatagaagaatgtggccgcgaagatctccgttagct | 1200 |
| QY | 1529 | gctggtcttcaatgctgtgtatatactcttggccggtgtgcttctcctaactatacaaca | 1588 |
| Db | 1201 | gctggtcttcaatgctgtgtatatactcttggccggtgtgcttctcctaactatacaaca | 1260 |
| QY | 1589 | ctggagagcgtttgtctctcttctgatttagcatcaacatctgabatgtctctcgtat | 1648 |
| Db | 1261 | ctggagagcgtttgtctctcttctgatttagcatcaacatctgabatgtctctcgtat | 1320 |
| QY | 1649 | gctgtatctctccctgggcatcaacaataatttgcacatctccagaaatggttgggcc | 1708 |
| Db | 1321 | gctgtatctctccctgggcatcaacaataatttgcacatctccagaaatggttgggcc | 1380 |
| QY | 1709 | gtcattcttaaaagtctgcaccctctgataaacatggttgggaatgtgcaaacggtttcat | 1768 |
| Db | 1381 | gtcattcttaaaagtctgcaccctctgataaacatggttgggaatgtgcaaacggtttcat | 1440 |
| QY | 1769 | atgctgctgctatlaaagtgttttggctgcatcttcttatacctatcttgcacaagtgtga | 1828 |

| | | | |
|----|------|---|------|
| Dd | 1441 | atgtctgtcgtcatuaatguttcttvtgtgcacattctcttaacatattcgcacaagvtgaa | 1500 |
| Qy | 1829 | gtacaaaacttgggtctctccaatgatcacacacagacacagacactgaaagagacaataaata | 1888 |
| Dd | 1501 | gtacaaaacttgggtctctccaatgatcacacatgacacagacacagacactgaaagagacaataaata | 1560 |
| Qy | 1889 | atccctgctctataatgatatatttatattcaatgtaacctcaagtgctctctgtatgt | 1944 |
| Dd | 1561 | atccctgctctataatgatatatttatattcaatgtaacctcaagtgctctctgtatgt | 1620 |
| Qy | 1949 | gttaagcattcatgtctcttttttaattgtactgttattagattttttaagggcttaatc | 2008 |
| Dd | 1621 | gttaagcattcatgtctcttttttaattgtactgttattagattttttaagggcttaatc | 1680 |
| Qy | 2009 | atgaaatactcatagtctgcagcaataataaatagtcaactgttctaatttatgataataatg | 2066 |
| Dd | 1681 | atgaaatactcatagtctgcagcaataataaatagtcaactgttctaatttatgataataatg | 1744 |
| Qy | 2069 | taagctagactctctactcttaagtgtccacatacctgcctgtcagtcgggcacaatgaata | 2128 |
| Dd | 1741 | taagctagactctctactcttaagtgtccacatacctgcctgtcagtcgggcacaatgaata | 1800 |
| Qy | 2129 | ggacaggtctgttgatttttttagggccatctactaaagggaatgagctgcaaacagactctct | 2188 |
| Dd | 1801 | ggacaggtctgttgatttttttagggccatctactaaagggaatgagctgcaaacagactctct | 1860 |
| Qy | 2189 | gataccctgtcttaattaaactatgatataatctctcaagttacgtatgaataaacactgttgt | 2248 |
| Dd | 1861 | gataccctgtcttaattaaactatgatataatctctcaagttacgtatgaataaacactgttgt | 1920 |
| Qy | 2249 | gttcaactctccataaaataatgttcagctctctctcgtcacactaagaccttaaacactttaga | 2308 |
| Dd | 1921 | gttcaactctccataaaataatgttcagctctctctcgtcacactaagaccttaaacactttaga | 1980 |
| Qy | 2309 | tctctgttgagactctccatacactgtataaatttcggcctgtgcacactgtgaactgaaaggaggt | 2366 |
| Dd | 1981 | tctctgttgagactctccatacactgtataaatttcggcctgtgcacactgtgaactgaaaggaggt | 2040 |
| Qy | 2369 | ggcccaagcagatctgcacaagcacatccctccctcgtgcttcaagggtcagaagtgcacagcgttat | 2428 |
| Dd | 2041 | ggcccaagcagatctgcacaagcacatccctccctcgtgcttcaagggtcagaagtgcacagcgttat | 2100 |
| Qy | 2429 | cagagtgcaagcatcccaagcccaagagccagtgctgactctctccgctgtgtgccttccctctga | 2488 |
| Dd | 2101 | cagagtgcaagcatcccaagcccaagagccagtgctgactctctccgctgtgtgtgccttccctctga | 2160 |
| Qy | 2489 | gggggtatccaatgtgtgataaagcccttgatgttggtgcaaaagacagtgagatccaactgcat | 2548 |
| Dd | 2161 | gggggtatccaatgtgtgataaagcccttgatgttggtgcaaaagacagtgagatccaactgcat | 2220 |
| Qy | 2549 | ggtcttgtataactctcaaaacttccctccccacagacagaaagaaataatgtgcgtgcatatga | 2608 |
| Dd | 2221 | ggtcttgtataactctcaaaacttccctccccacagacagaaagaaataatgtgcgtgcatatga | 2280 |
| Qy | 2609 | acctgtcaaaaagaaaatgtcgaaagcggccgggcacggtgtgctcatatgtctttaatcccaaga | 2668 |
| Dd | 2281 | acctgtcaaaaagaaaatgtcgaaagcggccgggcacggtgtgctcatatgtctttaatcccaaga | 2340 |
| Qy | 2669 | ctttgtgggggtctgaaggtgtgggcgaatacgaatgacagagatctcugaaaccaagccctgtgcagc | 2728 |
| Dd | 2341 | ctttgtgggggtctgaaggtgtgggcgaatacgaatgacagagatctcugaaaccaagccctgtgcagc | 2400 |
| Qy | 2729 | atgtgtgaaaccccatctctacttaaaaaatacaaaaaattagctgcgggcgtgtgtgaacgggc | 2788 |
| Dd | 2401 | atgtgtgaaaccccatctctacttaaaaaatacaaaaaattagctgcgggcgtgtgtgaacgggc | 2460 |
| Qy | 2789 | ctgttaatcccaatatactcagagagctgtgaggttaagaanaatcactctgaacctgtggaggtgga | 2848 |
| Dd | 2461 | ctgttaatcccaatatactcagagagctgtgaggttaagaanaatcactctgaacctgtggaggtgga | 2520 |
| Qy | 2849 | agttgtcagttgaaccaagaatacagccactgactccacgctctgggcgtatgtgagcgtgaactcc | 2908 |

Db 1201 gttatgatgatccctgctgctgtaacagctgctgacatattaaaggagcaaatgttttc 1260
 QY 1462 aaccttctgtctgcaggaatttttaagctatagagaattgttgaactgtgaattccct 1521
 Db 1261 aaccttaagtgctgcaggaatttttaagctatagagaattgttgaactgtgaattccct 1320
 QY 1522 ggtagctgctgcttcaattggtctgtgattatctcttgcgcgttcttcccaataatc 1581
 Db 1321 ggtagctgctgcttcaattggtctgtgattatctcttgcgcgttcttcccaataatc 1380
 QY 1582 aacacacatggaagcttttgcctctctgattatagacataacacatctgtatctctc 1641
 Db 1381 aacacacatggaagcttttgcctctctgattatagacataacacatctgtatctctc 1440
 QY 1642 ttctatgctgtatctctcctgcgggacacacacacatttcacatttcaggaaaggt 1701
 Db 1441 ttctatgctgtatctctcctgcgggacacacacacatttcacatttcaggaaaggt 1500
 QY 1702 tgggcgcgtcatctgtaaaaagctgacccctgataacactgttggagaatgcaaacgt 1761
 Db 1501 tgggcgcgtcatctgtaaaaagctgacccctgataacactgttggagaatgcaaacgt 1560
 QY 1762 gttctatattgtctgtctatataatgttttggcgcaatttcttaccattctgcaca 1821
 Db 1561 gttctatattgtctgtctatataatgttttggcgcaatttcttaccattctgcaca 1620
 QY 1822 aggtgaagtaacaaactgaggtctcctaattgacacacatgagacacagacatgaagaa 1881
 Db 1621 aggtgaagtaacaaactgaggtctcctaattgacacacatgagacacagacatgaagaa 1680
 QY 1882 ataaataatccctgcctctatthaatgtatatttattatcatgtaaacctcaaatgtccttc 1941
 Db 1681 ataaataatccctgcctctatthaatgtatatttattatcatgtaaacctcaaatgtccttc 1740
 QY 1942 tgatttgtaagaattctatgctcttttttttaattgtaactgtttagattttaagcc 2001
 Db 1741 tgatttgtaagaattctatgctcttttttttaattgtaactgtttagattttaagcc 1800
 QY 2002 tataatcatgaataatcatcatgattgacagaaataaataaactgtgttaattatga 2061
 Db 1801 tataatcatgaataatcatcatgattgacagaaataaataaactgtgttaattatga 1860
 QY 2062 taataatgaactgagacacttctactttaggttgcacataactgcctgctagtcgggca 2121
 Db 1861 taataatgaactgagacacttctactttaggttgcacataactgcctgctagtcgggca 1920
 QY 2122 tgaagttaggaagcttctgttatttttttttaagggccatactaaaggaaatgagtaaac 2181
 Db 1921 tgaagttaggaagcttctgttatttttttttaagggccatactaaaggaaatgagtaaac 1980
 QY 2182 accctcgtatcccttgccttaataaactagatgataattctcagctgacatgataaac 2241
 Db 1981 accctcgtatcccttgccttaataaactagatgataattctcagctgacatgataaac 2040
 QY 2242 tgttgtgttcaacttctccataaanaattgtcagctctctctgacacttagaactcaaac 2301
 Db 2041 tgttgtgttcaacttctccataaanaattgtcagctctctctgacacttagaactcaaac 2100
 QY 2302 tttagcatctctgtgagagcgccacacacatgataatttgcgcgggaactgagctgag 2361
 Db 2101 tttagcatctctgtgagagcgccacacacatgataatttgcgcgggaactgagctgag 2160
 QY 2362 ggaatgagcccaagcagctgcagcaactccctccctgcgttcaggggtcagagcgccag 2421
 Db 2161 ggaatgagcccaagcagctgcagcaactccctccctgcgttcaggggtcagagcgccag 2220
 QY 2422 cgtttatcagagcagcagatcccaagccagagccagtgctgactcttgcgtgtgtcctt 2481
 Db 2221 cgtttatcagagcagcagatcccaagccagagccagtgctgactcttgcgtgtgtcctt 2280
 QY 2482 cctctgaggggctatcaaatgtgtatgataaagccctgagtaggaagcagtgagatcca 2541
 Db 2281 cctctgaggggctatcaaatgtgtatgataaagccctgagtaggaagcagtgagatcca 2340

QY 2542 ctgtcatggtcttgatatacatctccaaacttcccttcccaagcacaggaatatgtgtg 2601
 Db 2341 ctgtcatggtcttgatatacatctccaaacttcccttcccaagcacaggaatatgtgtg 2400
 QY 2602 gcatgcaacctgcaaaaagaataatgcaagcgcgggcaacgtgtgctcatgctgta 2661
 Db 2401 gcatgcaacctgcaaaaagaataatgcaagcgcgggcaacgtgtgctcatgctgta 2460
 QY 2662 cccagcaacttggggggtgtgaggtgtggcgaaatcatgtgatacagagttcgagaccac 2721
 Db 2461 cccagcaacttggggggtgtgaggtgtggcgaaatcatgtgatacagagttcgagaccac 2520
 QY 2722 ggcagcatgtgcaaaccccatctctactaaataacaaaaaatgaactgtggcggtgta 2781
 Db 2521 ggcagcatgtgcaaaccccatctctactaaataacaaaaaatgaactgtggcggtgta 2580
 QY 2782 cgggcgcctgtaatccagatactcaagagcgtgagtaggaagaatacactggaacctgg 2841
 Db 2581 cgggcgcctgtaatccagatactcaagagcgtgagtaggaagaatacactggaacctgg 2640
 QY 2842 aggtggaagttgcaagtgaaacaaagatacagccactgacacagcctggcgatggagcg 2901
 Db 2641 agatggaagttgcaagtgaaacaaagatacagccactgacacagcctggcgatggagcg 2700
 QY 2902 agactccaactc 2913
 Db 2701 agactccaactc 2712

RESULT 5
 AAF55900
 ID AAF55900 standard; DNA; 2512 BP.
 XX
 AC AAF55900;
 DT 18-APR-2001 (first entry)
 XX
 DE Human AST coding sequence.
 XX
 KW Human; AST; nootropic; immunotropic; gene therapy; Salla disease;
 KW anion and sugar transporter; anion-cation symporter;
 KW Salla acid transporter; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1069184-A1.
 PD 17-JAN-2001.
 XX
 PF 16-JUL-1999; 99EP-0202341.
 PR 16-JUL-1999; 99EP-0202341.
 XX
 PA (ALKU) AKZO NOBEL NV.
 XX
 DR WPI: 2001-193090/20.
 DR P-PSDB; AAB66967.
 XX
 PT New human transporter gene implicated in Salla disease and lysosomal
 PT Salla acid transport, useful in assays for identifying new drugs, or
 PT diagnosing Salla acid transport defects related to mutations in the
 PT transporter gene -
 XX
 PS Claim 2; Page 12-13; 20pp; English.
 XX
 CC The present sequence is the coding sequence for human Anion and Sugar
 CC Transporter (AST) protein. AST has significant homology with several
 CC members of the anion-cation symporter (ACS) family of transporters. AST
 CC is implicated in Salla disease, and is useful in screening assays for
 CC identifying new drugs. Compounds identified via AST screening are useful
 CC for preparing a pharmaceutical suitable as an activator or inhibitor of a
 CC Salla acid transporter protein. The pharmaceutical may be used in Salla

CC acid associated diseases and CNS/immune related disorders.
XX
SQ Sequence 2512 BP; 612 A; 585 C; 566 G; 749 T; 0 other;

| | | | | |
|-----------------------|-----------------|---------------|------------|--------------|
| Query Match | 80.5%; | Score 2359.4; | DB 22; | Length 2512; |
| Best Local Similarity | 97.28; | Pred. No. 0; | | |
| Matches 2437; | Conservative 0; | Mismatches 6; | Indels 64; | Gaps 1; |

[illegible]

| | | | |
|----|------|---|-------|
| QY | 1081 | tttctgtactatttggaaatttttggtctcttctgtgtacctggttaagttagtcacaacc | 1140 |
| DB | 966 | ttttctgtactatttggaaatttttggtctcttctgtgtacctggttaagttagtcacaacc | 10255 |
| QY | 1141 | acaaacaacaaaggaaatttcccaatttgaagaagaaatcatcttccatcaataaga | 1200 |
| DB | 1026 | acaaacaacaaaggaaatttcccaatttgaagaagaaatcatcttccatcaataaga | 10855 |
| QY | 1201 | tcaagttcttccacaagaagtcagtcggttggtgacacattttaaattccctccgcaattg | 1260 |
| DB | 1086 | tcaagttcttccacaagaagtcagtcggttggtgacacattttaaattccctccgcaattg | 1145 |
| QY | 1261 | ggctatcgtagtgcacacttttcccttaacaactggaacttttaacttaattgacaattc | 1320 |
| DB | 1146 | ggctatcgtagtgcacacttttcccttaacaactggaacttttaacttaattgacaattc | 1205 |
| QY | 1321 | ggctactatrttgaagaagttcccaagtttcaatttgaagaagatgggtttttatctc | 1380 |
| DB | 1206 | ggctactatrttgaagaagttcccaagtttcaatttgaagaagatgggtttttatctc | 1265 |
| QY | 1381 | attgcttattttaagcctctggttatgtatgaagaccgtctgtgcaagctgcgcgaacatt | 1440 |
| DB | 1266 | attgcttattttaagcctctggttatgtatgaagaccgtctgtgcaagctgcgcgaacatt | 1325 |
| QY | 1441 | aaggcgcaaaatlgaaatttttccaacttaltgtgtcgcagaacattttaagcctatagaat | 1500 |
| DB | 1326 | aaggcgcaaaatlgaaatttttccaacttaltgtgtcgcagaacattttaagcctatagaat | 1385 |
| QY | 1501 | gatttgaccctgcagatcttcctgtgtagctgcctgtgctcaattgctgtgatatcttctg | 1560 |
| DB | 1386 | gatttgaccctgcagatcttcctgtgtagctgcctgtgctcaattgctgtgatatcttctg | 1445 |
| QY | 1561 | cggtgcttccctaactatcatcaacaacaactgggaagcctttgctctcttgatattagat | 1620 |
| DB | 1446 | cggtgcttccctaactatcatcaacaacaactgggaagcctttgctctcttgatattagat | 1505 |
| QY | 1621 | caaccatctgtgatatgtctcctctgtaattgcgtgtatccctcctgggcacacaatacatt | 1680 |
| DB | 1506 | caaccatctgtgatatgtctcctctgtaattgcgtgtatccctcctgggcacacaatacatt | 1565 |
| QY | 1681 | tgccaactatccagaagatggtttggcccgctcatctgtcaaaatctcgaccctgataaac | 1740 |
| DB | 1566 | tgccaactatccagaagatggtttggcccgctcatctgtcaaaatctcgaccctgataaac | 1625 |
| QY | 1741 | tgttggagaagtcggaacccgtgtctctcatatattgtctgtgcatatgttttttggtcgat | 1800 |
| DB | 1626 | tgttggagaagtcggaacccgtgtctctcatatattgtctgtgcatatgttttttggtcgat | 1685 |
| QY | 1801 | tttctttaaactattcgcgaagtcgaagtacaaaaactgggctctcaatgatcacacatg | 1860 |
| DB | 1686 | tttctttaaactattcgcgaagtcgaagtacaaaaactgggctctcaatgatcacacatg | 1745 |
| QY | 1861 | acacagacactcgaaggaacaaataaataatcctgtcctctataatgtatatttattatc | 1920 |
| DB | 1746 | acacagacactcgaaggaacaaataaataatcctgtcctctataatgtatatttattatc | 1805 |
| QY | 1921 | atgttaacctcaaatgtgctctctgtattgttggaagcttcatgtcttttttaattgac | 1980 |
| DB | 1806 | atgttaacctcaaatgtgctctctgtattgttggaagcttcatgtcttttttaattgac | 1865 |
| QY | 1981 | ttgtattagaattttttaagcctataatcaatgaatatacactatgttgcagaataataa | 2040 |
| DB | 1866 | ttgtattagaattttttaagcctataatcaatgaatatacactatgttgcagaataataa | 1925 |
| QY | 2041 | tgaactgtgtttaaattatgaataatgtgaagctaggaacttacttaagtgtcaacatc | 2100 |
| DB | 1926 | tgaactgtgtttaaattatgaataatgtgaagctaggaacttacttaagtgtcaacatc | 1985 |
| QY | 2101 | ctgcgcctcgtctgctgggaataatgaataagaacagttctgtgtattttttaaggccatct | 2160 |
| DB | 1986 | ctgcgcctcgtctgctgggaataatgaataagaacagttctgtgtattttttaaggccatct | 2045 |
| QY | 2161 | aaagggaatagctcgaaacagcactctctgataacacttcttgtaatttaaacctgacatcat | 2220 |

|||||
Db 2046 aaaggaatgagctgaaacagaccctctgataccttgccttaataactagatgataat 2105
Qy 2221 tctaaagtaactgataaacaacctgtgtgttgcactcttcctccataaataatgtaagctctc 2280
Db 2106 tctaaagtaactgataaacaacctgtgtgttgcactcttcctccataaataatgtaagctctc 2165
Qy 2281 tctgacacttagacactcaacttagatcactctgtgtgagctgacatccatcactgataatt 2340
Db 2166 tctgacacttagacactcaacttagatcactctgtgtgagctgacatccatcactgataatt 2225
Qy 2341 cgctctgcaactgtgactgagggaggtgtgcccaagcagctgtgccaagcactccctctg 2400
Db 2226 cgctctgcaactgtgactgagggaggtgtgcccaagcagctgtgccaagcactccctctg 2285
Qy 2401 cttaaggtgcaaggtgcccaagcgtttatcaagggcaagcattcccaagcccaagcaaggttc 2460
Db 2286 cttaaggtgcaaggtgcccaagcgtttatcaagggcaagcattcccaagcccaagcaaggttc 2345
Qy 2461 gactcttcgctgtgtgcttctcctctctgaggggcctatcattgtgtagataaagccctgagt 2520
Db 2346 gactcttcgctgtgtgcttctcctctctgaggggcctatcattgtgtagataaagccctgagt 2405
Qy 2521 aggaagaagaagctgagatccactgtctatgttctgtatatacattcccaacttccctcc 2580
Db 2406 aggaagaagaagctgagatccactgtctatgttctgtatatacattcccaacttccctcc 2465
Qy 2581 agccaagaagaatattgtctgtgcatgcaacgtgcaagaagaataatgc 2627
Db 2466 agccaagaagaatattgtctgtgcatgcaacgtgcaagaagaataatgc 2512
RESULT 6
AAH79234
ID AAH79234 standard; cDNA; 2670 BP.
XX
AC AAH79234;
XX
DT 20-NOV-2001 (first entry)
XX
DE Human sodium dependent phosphate co-transfer protein 35 cDNA.
XX
KW Human; sodium dependent phosphate co-transfer protein 35;
KW hypophosphaturia; hypercalcaemia; hypophosphataemic rickets;
KW nephritis; gene therapy; ss.
XX
OS Homo sapiens.
PN CN1298882-A.
XX
PD 13-JUN-2001.
XX
PE 06-DEC-1999; 99CN-0124217.
XX
PR 06-DEC-1999; 99CN-0124217.
XX
PA (SHEN-) SHENGYUAN GENE DEV CO LTD SHANGHAI.
XX
PI Mao Y, Xie Y;
XX
XX WPI, 2001-503367/56.
DR P-PSDB; AAG65238.
XX
PT Human Na-dependent phosphate cotransporter 35 and its coding sequence -
XX
PS Claim 6; Page 19-20(Disclosure); 28pp; Chinese.
XX
CC The present invention provides the protein and coding sequences of human
CC sodium dependent phosphate co-transfer protein 35. The sequences can be
CC used in the treatment of hypophosphaturia, hypercalcaemia,
CC hypophosphataemic rickets and nephritis. The present sequence is the
CC coding sequence of the invention.
XX

SQ Sequence 2670 BP; 706 A; 550 C; 552 G; 862 T; 0 other;
Query Match 68.9%; Score 2018.6; DB 22; Length 2670;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2021; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 906 gaagagggtgtatacattccacgcatgcatgcatgtgtgtcttctgtggtccccccttg 965
Db 1 gaagagggtgtatacattccacgcatgcatgcatgtgtgtcttctgtggtccccccttg 60
Qy 966 aaagaagaacaactcttagcatttcgtatgcagagacagcagcttggagcaatattctc 1025
Db 61 aaagaagaacaactcttagcatttcgtatgcagagacagcagcttggagcaatattctc 120
Qy 1026 ttccctcttgtaataatttgcctacatgataatgattgatttgccttcttcttcttctt 1085
Db 121 ttccctcttgtaataatttgcctacatgataatgattgatttgccttcttcttcttctt 180
Qy 1086 gtaactatgataatttctgttctcttctgtgactgtgttagttagtacacacacaa 1145
Db 181 gtaactatgataatttctgttctcttctgtgactgtgttagttagtacacacacaa 240
Qy 1146 aacaaagaagaattcccatatgaaagaataacattcttcaatcaataagaatacagc 1205
Db 241 aacaaagaagaattcccatatgaaagaataacattcttcaatcaataagaatacagc 300
Qy 1206 ttcttcaagaagaattcccatatgaaagaataacattcttcaatcaataagaatacagc 1265
Db 301 ttcttcaagaagaattcccatatgaaagaataacattcttcaatcaataagaatacagc 360
Qy 1266 tctgtagttgacacatttcttcaactgacatttcaactatgataattatgataattgcta 1325
Db 361 tctgtagttgacacatttcttcaactgacatttcaactatgataattatgataattgcta 420
Qy 1326 ctataatgaagaagatccctaaagtccaatgttcaagaagaatggttttcttcaatgc 1385
Db 421 ctataatgaagaagatccctaaagtccaatgttcaagaagaatggttttcttcaatgc 480
Qy 1386 ctataatgaagaagatccctaaagtccaatgttcaagaagaatggttttcttcaatgc 1445
Db 481 ctataatgaagaagatccctaaagtccaatgttcaagaagaatggttttcttcaatgc 540
Qy 1446 caaatatgaatttcttcaactatgtgtctgcagaaattttaaactatagataatgata 1505
Db 541 caaatatgaatttcttcaactatgtgtctgcagaaattttaaactatagataatgata 600
Qy 1506 gacctgacatattccctgtgtagctgtgcttcaatgtgctgtgataattcttctgcccgtg 1565
Db 601 gacctgacatattccctgtgtagctgtgcttcaatgtgctgtgataattcttctgcccgtg 660
Qy 1566 ctcttccaaactatatacaacaacacgtggagccttgccttcttcttcttcttcttctt 1625
Db 661 ctcttccaaactatatacaacaacacgtggagccttgccttcttcttcttcttcttctt 720
Qy 1626 atctgatatgtctcctctgtatgctgtatctcctctggtgacacaaataactgtgcca 1685
Db 721 atctgatatgtctcctctgtatgctgtatctcctctggtgacacaaataactgtgcca 780
Qy 1686 ctattccagaagaatggttggcccgctcaatgtctaaagctgacccctgataacactgttg 1745
Db 781 ctattccagaagaatggttggcccgctcaatgtctaaagctgacccctgataacactgttg 840
Qy 1746 gagaatggcaaacgcgttctatattgtgctgtcctttaaattgttttgggcaatttct 1805
Db 841 gagaatggcaaacgcgttctatattgtgctgtcctttaaattgttttgggcaatttct 900
Qy 1806 ttacacattcgcgcaagaagtgaagtaacaacaacggtgctccaatgatacaccagagaca 1865
Db 901 ttacacattcgcgcaagaagtgaagtaacaacaacggtgctccaatgatacaccagagaca 960
Qy 1866 gacctgaaggaacaataataatccctgctctatattatatttattatcatgta 1925
Db 960 gacctgaaggaacaataataatccctgctctatattatatttattatcatgta 1925

| | | | |
|----|------|---|------|
| Dh | 961 | gacctgaaggaaccaataataatccctgcctctatattgaatgtattttattatcahyta | 1020 |
| Qy | 1926 | acctcaagctgccttcctgtatctgtgaagcattctatctgtctttttaaattgtaactgta | 1985 |
| Dh | 1021 | acctcaagctgccttcctgtatctgtgaagcattctatctgtctttttaaattgtaactgta | 1080 |
| Qy | 1966 | ttgaattttaaggcctatataatcatgaataatcatcactagtctgcagaaataataatgaa | 2045 |
| Dh | 1081 | ttgaattttaaggcctatataatcatgaataatcatcactagtctgcagaaataataatgaa | 1140 |
| Qy | 2046 | tgctttaattatgaataatctgtgaagcaggactctcacttaagtttgtaactatcctgc | 2105 |
| Dh | 1141 | tgctttaattatgaataatctgtgaagcaggactctcacttaagtttgtaactatcctgc | 1200 |
| Qy | 2106 | tgctagctgcggcaacatgaagtgaagcagctctctgtgatcttttaaggccataaagg | 2165 |
| Dh | 1201 | tgctagctgcggcaacatgaagtgaagcagctctctgtgatcttttaaggccataaagg | 1260 |
| Qy | 2166 | gaatgaagctgaagaacagacccctgcctgactctgtgtaatttaactagatgatacttca | 2225 |
| Dh | 1261 | gaatgaagctgaagaacagacccctgcctgactctgtgtaatttaactagatgatacttca | 1320 |
| Qy | 2226 | ggtactatgaacaacactgtctgtctgtctacactcttcctcataaanaatgtcagctcctc | 2285 |
| Dh | 1321 | ggtactatgaacaacactgtctgtctgtctacactcttcctcataaanaatgtcagctcctc | 1380 |
| Qy | 2286 | cacttagaacctcaaaacttagacatctctctgtgaagctgcacatccactgtataattgcct | 2345 |
| Dh | 1381 | cacttagaacctcaaaacttagacatctctctgtgaagctgcacatccactgtataattgcct | 1440 |
| Qy | 2346 | ggcaacctggaactagaagggaatgtgcccagacggcctgcagcaaacctccctctgctca | 2405 |
| Dh | 1441 | ggcaacctggaactagaagggaatgtgcccagacggcctgcagcaaacctccctctgctca | 1500 |
| Qy | 2406 | gggtcagaaagtgcaccaagcgtttatcagaaggcagcatccaaagcccaaggcagtgctgaatc | 2465 |
| Dh | 1501 | gggtcagaaagtgcaccaagcgtttatcagaaggcagcatccaaagcccaaggcagtgctgaatc | 1560 |
| Qy | 2466 | ttcgccgcgtggtccttccctctctgaagggtcatcaaatgtctagaataaggccgtgaatgca | 2525 |
| Dh | 1561 | ttcgccgcgtggtccttccctctctgaagggtcatcaaatgtctagaataaggccgtgaatgca | 1620 |
| Qy | 2526 | agagcagctgaagatccacatgctatctgtcttgatatacatccccaactttccctcccaagac | 2585 |
| Dh | 1621 | agagcagctgaagatccacatgctatctgtcttgatatacatccccaactttccctcccaagac | 1680 |
| Qy | 2586 | agaggaatatcttgctgtgcatatgcacactctgcaaaaaagaanaatctcgaaagcggccggacagct | 2645 |
| Dh | 1681 | agaggaatatcttgctgtgcatatgcacactctgcaaaaaagaanaatctcgaaagcggccggacagct | 1740 |
| Qy | 2646 | ggcccaatgcgtctgaatccccaagcactcttgaggggcgtgaaggttgccgaatcatgaagatcagg | 2705 |
| Dh | 1741 | ggcccaatgcgtctgaatccccaagcactcttgaggggcgtgaaggttgccgaatcatgaagatcagg | 1800 |
| Qy | 2706 | agttctgaagccagcctgtgcccagcatatgtgaaaaccccatctctactataaataatacaaaaat | 2765 |
| Dh | 1801 | agttctgaagccagcctgtgcccagcatatgtgaaaaccccatctctactataaataatacaaaaat | 1860 |
| Qy | 2766 | taagctggggtgctgtgagcgggctgttaatctccagatctcaagaggtctgaggtgaagaga | 2825 |
| Dh | 1861 | taagctggggtgctgtgagcgggctgttaatctccagatctcaagaggtctgaggtgaagaga | 1920 |
| Qy | 2826 | atcaactgaaacccgggaaggtgtgaagattctgagcgaacccaagatcaacgcaactgcgactccag | 2885 |
| Dh | 1921 | atcaactgaaacccgggaaggtgtgaagattctgagcgaacccaagatcaacgcaactgcgactccag | 1980 |
| Qy | 2886 | ctctgggcgtatgagcgagagactccaactcaaaaaaaaaaaaaa 2930 | |
| Dh | 1981 | ctctgggcgtatgagcgagagactccaactcaaaaaaaaaaaaaa 2025 | |

RESULT 7
AA250876

| | | |
|----------------------------|--|--------------------------------------|
| ID | AAZ50876 | standard; cDNA; 2844 BP. |
| XX | AAZ50876; | |
| AC | | |
| XX | | |
| DT | 31-MAY-2000 | (first entry) |
| XX | | |
| DE | Sheep GBS toxin receptor (SP55) cDNA. | |
| XX | | |
| KW | Sheep GBS toxin receptor; group B beta-hemolytic streptococci; SP55; | |
| KM | pathological vascularisation; cancer metastases; angiogenesis; | |
| KV | neovascularisation; reperfusion injury; scarring; keloid; | |
| KX | chronic inflammation; rheumatoid arthritis; psoriasis; neural injury; | |
| KW | endothelial cell proliferation; antibacterial; anticancer; | |
| XX | anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss. | |
| OS | Ovis sp. | |
| FH | Key | Location/Qualifiers |
| FT | CDS | 84..1571 |
| FT | | /tag= a |
| FT | | /product= "Sheep GBS toxin receptor" |
| PX | WO200005375-A1. | |
| XX | | |
| PD | 03-FEB-2000. | |
| PF | 22-JUL-1999; | 99WO-US16676. |
| XX | | |
| PR | 22-JUL-1998; | 98US-0093843. |
| XX | | |
| PA | (UYVA-) UNIV VANDERBILT. | |
| XX | | |
| PI | Hellergvist CG, Fu C; | |
| XX | | |
| DR | WPI; 2000-205377/18. | |
| P-PSDB: | AAV45088. | |
| PT | New polynucleotide encoding mammalian receptor for streptococcus toxin, | |
| XX | useful for diagnosis and treatment of, e.g. pneumonia in neonates - | |
| PS | Claim 3; Page 83-86; 109pp; English. | |
| XX | | |
| CC | The present cDNA sequence encodes partial sheep GBS (group B beta | |
| CC | -hemolytic Streptococci) toxin receptor (SP55). This sequence was cloned | |
| CC | using a primary culture of sheep lung endothelial cells. Expression | |
| CC | vectors comprising this cDNA can be transformed into host cells to | |
| CC | express GBS toxin receptor and its fragments. Detecting the receptor in | |
| CC | tissues is used to diagnose pathological vascularisation, e.g. for | |
| CC | detecting cancer metastases. GBS toxin receptors are useful for treating | |
| CC | conditions associated with pathological angiogenesis or | |
| CC | neovascularisation (specifically cancer, reperfusion injury, scarring | |
| CC | during wound healing, keloids, chronic inflammation (rheumatoid | |
| CC | arthritis or psoriasis) or neural injury), and to raise specific | |
| CC | antibodies used for treating early onset disease. Inhibitors of this | |
| CC | receptor are useful for treating pathological or hypoxia-induced | |
| CC | endothelial cell proliferation and migration. | |
| XX | | |
| SQ | Sequence 2844 BP; 745 A; 623 C; 594 G; 882 T; 0 other: | |
| <hr/> | | |
| Query Match | 53.6%; Score 1570.8; DB 21; Length 2844; | |
| Best Local Similarity | 82.0%; Pred. No. 0; | |
| Matches 1916; Conservative | 0; Mismatches 387; Indels 33; Gaps | 8 |
| OY | 309 ggcgccgggagcgttggcgctgctgcgcgcgcgtcccttccttgccacgagtggcgatcac | 368 |
| DB | | 64 |
| OY | 5 ggaggcgaggggcgttcgctgcgcgcgcgtcccttccttcgacgagcaggtttgcgc | 64 |
| OY | 369 ct--gtcacgtaagcgtatgatgactcgcgttcgacgaccgtgcccgaacatggca | 426 |
| DB | | 124 |
| OY | 65 gtacctccctcgtaaagcatatgaattccccggttttcgacttaagccccgaagcgaca | 124 |
| OY | 427 ggaagacagaagccacacgctctcttaccagggcgcccaacagggccgaagcgttcagtl | 486 |

Db 1145 aagaatgaattttcaactctgtgttcgaagatttttagccttaagatgattcg 1204
QY 1507 accgcgagatctcctgttagctgtgccttcacatgcgtgatattcttggccgttg 1566
Db 1205 accgcgagatctcctgttagctgtgccttcacatgcgtgatattcttggccgttg 1264
QY 1567 ttctctaactatcaacaacacactggagccttctgtcctcttgatttagatcaaca 1626
Db 1265 attcctaaccatcaacaacacactggagccttctgtcctcttgatttagatcaaca 1324
QY 1627 tctgtgatatgtctcctcgtgtgtgtgtatcctcctgggcatcaaatatctggcac 1686
Db 1325 tctgtgatatgtctcctcgtgtgtgtgtatcctcctgggcatcaaatatctggcac 1384
QY 1687 tattccaggaatgtgtggccgcgtcatgttaaaagtctgaacccctgatacactgttg 1746
Db 1385 tattccaggaatgtgtggccgcgtcatgttaaaagtctgaacccctgatacactgttg 1444
QY 1747 agaatggcaaacgltgtctatattgtcgtcgtatattgttttgggcatcttct 1806
Db 1445 agaatggcaaacgltgtctatattgtcgtcgtatattgttttgggcatcttct 1504
QY 1807 tacacttgcgcaaaagtgaagtaaacaaactggcctcctaataatatacattgacacag 1866
Db 1505 cacactatcgcgaagtgtaagtgcaaaactggccatcagtgatccaaagaacacag 1564
QY 1867 acactgaaggaaccaataataaactcgtcctctataatgtattctatcatgttaa 1926
Db 1565 aaactgaaggaaccaataataaactcgtcctctataatgtattcttattatcatgttaa 1624
QY 1927 cctcaaatgtcctt-----ctgtatgtgtgaagcatctcatgt-----ctttttt 1972
Db 1625 cctcaaatgtccttctgtatattttaaigtgaagcaatcatatatacaagaataaattgtac 1684
QY 1973 aattgtactgtatatttaagtcgctataatcgtataatcgtatcgttgacgaa 2032
Db 1685 tagaanaattgtgtctgaatctgtgaagctgtgaatcgtgaatgtcactgttgccatat 1744
QY 2033 taataaataatgaactgttcttaataataataatgaatgaagcctaagacttcaacttagt 2092
Db 1745 aagcaaaatttagctatttcttaattattattataccctgttgctgaacttaaatcagg 1804
QY 2093 tcaatatcctgcctcgtcagtcgcgcaacatgtaagtaggaagctcgtgtgatttttaag 2152
Db 1805 tcaatatcctgcctcgtcagtcgcgcaacatgtaagtaggaagctcgtgtgatttttaag 1863
QY 2153 ggcataactaagggaatgaactgaacaacacctcctgatcccttggtttaataaactag 2212
Db 1864 accataactaagggaatgaactgaacaacacctcctgatcccttggtttaataaactag 1923
QY 2213 atgataatctcagtgatgataaacaactgtgtgtgtcacttccctataaana----- 2267
Db 1924 ataataatctcagtgatgataaacaactgtgtgtgtgtcacttccctataaana----- 1983
QY 2268 -atgtgtagctctctcctgacacttaagaccctcaaaacttaacatctcgtgtgagctgccat 2326
Db 1984 tgtatataagaatccctgcaatgtgaagctccaacttaagcctcctccacaggaagctggcag 2043
QY 2327 ccaactgataatctcgtcgtgcaactgagtgagtgagtggtgtcccgagggagcctcaag 2386
Db 2044 ccaactgataatctcgtcgtgcaactgagtgagtgagtggtgtcccgagggagcctcaag 2103
QY 2387 cactccctcctcgtcgtcagtgatcagagtgccagcgtttatcaagaagcagcatccagc 2446
Db 2104 tgtccctcctcgtcgtcagtgatcagagtgccagcgtttatcaagaagcagcatccagc 2156
QY 2447 ccagagcagcgtgtcagactcctcgtgtgtgtccttccctcgtgagggagcattatgtgtag 2506
Db 2157 ccagagcagcgtgtcagactcctcgtgtgtgtgtccttccctcgtgagggagcattatgtgtg 2214
QY 2507 ataaagccctgagtaggaagcagtaggaatcgaactcgtcatgtgtctgtataacactca 2566
Db 2215 atgaagccctgagtaggaagcagtaggaatcgaactcgtcatgtgtgtctgtataacactca 2273

QY 2567 aacttccctccagcagagaatattgctgtgcgtgacatgaacttcaaaagaa 2622
Db 2274 tcccttccctccagcagacactgtgagtagtgcctgtgcatgttaacttcaaaagaa 2329

RESULT 9

AA158115
ID AA158115 standard; cDNA; 1488 BP.

AC AA158115;

DE 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 318.

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

KW leukemia; ss.

XX Homo sapiens.

OS Homo sapiens.

PN WO200153312-A1.

XX 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI: 2001-442253/47.

DR P-PSDB; AAM38959.

PT Novel nucleic acids and polypeptides, useful for treating disorders

PS such as central nervous system injuries -

XX Claim 1; SEQ ID NO 318; 10078bp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and

CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: immune system suppression,

CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukaemias and

CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed

CC specification.

XX Sequence 1488 BP; 368 A; 318 C; 323 G; 479 T; 0 other;

| | | | | |
|-----------------------|--------------|---------------|------------|----------------------|
| Query Match | 50.7%; | Score 1484.8; | DB 22; | Length 1488; |
| Best Local Similarity | 99.9%; | Pred. No. 0; | | |
| Matches 1486; | Conservative | 0; | Mismatches | 2; Indels 0; Gaps 0; |

| | | | |
|----|------|---|------|
| QY | 386 | atgaggtctccggttcgagagccgctgcgcgcggagaaagatgagcgagaaagacagcgagccgacg | 445 |
| Db | 1 | atgaggtctccggttcgagagccgctgcgcgcggagaaagatgagcgagaaagacagcgagccgacg | 60 |
| QY | 446 | cctctctacccggcgcccccacggcgccgaagccgcctccagctgtgctgctctgctgttac | 505 |
| Db | 61 | cctctctacccggcgcccccacggcgccgaagccgcctccagctgtgctgctctgctgttac | 120 |
| QY | 506 | aactttagcaatttggcctttttgtttcttcctcatgtgtatgcatatcaatgtatgaactcg | 565 |
| Db | 121 | aactttagcaatttggcctttttgtttcttcctcatgtgtatgcatatcaatgtatgaactcg | 180 |
| QY | 566 | agtgctgcttaagtgaatgatagtatgataatcaaacacttttagaagataatagaacttcc | 625 |
| Db | 181 | agtgctgcttaagtgaatgatagtatgataatcaaacacttttagaagataatagaacttcc | 240 |
| QY | 626 | aagcggtccagagcatcttcgtcccaataaaagttcatcatcaacaacgggtgaag | 685 |
| Db | 241 | aagcggtccagagcatcttcgtcccaataaaagttcatcatcaacaacgggtgaag | 300 |
| QY | 686 | taccaatggaatgcagaaactcaagatgatatctcgcttcttttatagtctacac | 745 |
| Db | 301 | taccaatggaatgcagaaactcaagatgatatctcgcttcttttatagtctacac | 360 |
| QY | 746 | atccacacagattccctggagagataatgttcgacagaaataagggggaaaaagttgttata | 805 |
| Db | 361 | atccacacagattccctggagagataatgttcgacagaaataagggggaaaaagttgttata | 420 |
| QY | 806 | tttgagatccttgagcaatgtgtcttcctcacccgttttacctcccaattgtgtcagattaga | 865 |
| Db | 421 | tttgagatccttgagcaatgtgtcttcctcacccgttttacctcccaattgtgtcagattaga | 480 |
| QY | 866 | gttggaccactcaattgtaacctcaagacacataagaagactagaagaggtgtatcatattcca | 925 |
| Db | 481 | gttggaccactcaattgtaacctcaagacacataagaagactagaagaggtgtatcatattcca | 540 |
| QY | 926 | gcacatgcatgcacatgtgtcttcttcggctccccccttgaaagaagcaaatcttcagc | 985 |
| Db | 541 | gcacatgcatgcacatgtgtcttcttcggctccccccttgaaagaagcaaatcttcagc | 600 |
| QY | 986 | atttcgatacgagagacacagcctctggagacgtaattctctctcctctcttctgnaataat | 1045 |
| Db | 601 | atttcgatacgagagacacagcctctggagacgtaattctctctcctctcttctgnaataat | 660 |
| QY | 1046 | tgctactatataatgtgacttatgtctctactatttttggfractatgtgaatatattgg | 1105 |
| Db | 661 | tgctactatataatgtgacttatgtctctactatttttggfractatgtgaatatattgg | 720 |
| QY | 1106 | ttctctttcggatcctggttagttagttagtgacacccacaaacacaaagaaatttccat | 1165 |
| Db | 721 | ttctctttcggatcctggttagttagttagtgacacccacaaacacaaagaaatttccat | 780 |
| QY | 1166 | tatgaaaaagaatacatcttcttcatcaataagaanaatcaagcttcttccaagaagtcaagt | 1225 |
| Db | 781 | tatgaaaaagaatacatcttcttcatcaataagaanaatcaagcttcttccaagaagtcaagt | 840 |
| QY | 1226 | ccggtgggtaccacattttaaaatcccccgcacacttgggcatagtatgttgcacacttct | 1285 |
| Db | 841 | ccggtgggtaccacattttaaaatcccccgcacacttgggcatagtatgttgcacacacttct | 900 |
| QY | 1286 | tacaacacggacttttataactatgtacatatatgtccactactaataatgaagagatactta | 1345 |
| Db | 901 | tacaacacggacttttataactatgtacatatatgtccactactaataatgaagagatactta | 960 |
| QY | 1346 | aggttcaaatgttcaagaaatgggttttatacttcatatgtccttatattggcctctgtgta | 1405 |
| Db | 961 | aggttcaaatgttcaagaaatgggttttatacttcatatgtccttatattggcctctgtgta | 1020 |

| QY | 1406 | tgtatgatccgtctctggtccaactcgtctgccaatttaaggccaatggaatttttcaact | 1465 |
|-----------|---|---|------|
| QY | 1406 | tgtatgatccgtctctggtccaactcgtctgccaatttaaggccaatggaatttttcaact | 1465 |
| Db | 1021 | tgtatgatccgtctcgtgtcaaacctcgtctgccaatttaaggccaatggaatttttcaact | 1080 |
| QY | 1466 | ttaatgtttcgcaagaatttttagtccttaagaaatgattggaccgcgcgaattctcgtga | 1525 |
| Db | 1081 | ttaatgtttcgcaagaatttttagtccttaagaaatgattggaccgcgcgaattctcgtga | 1140 |
| QY | 1526 | gctctgctgcttcattgctgctgattatctctttggccgttgccttctcctaactaataca | 1585 |
| Db | 1141 | gctctgctgcttcattgctgctgattatctctttggccgttgccttctcctaactaataca | 1200 |
| QY | 1586 | aaactggaggagcttttgcctcttcctgatttagcatcaaccatctgatatctccttcg | 1645 |
| Db | 1201 | aaactggaggagcttttgcctcttcctgatttagcatcaaccatctgatatctccttcg | 1260 |
| QY | 1646 | tatgtctgatatcctctctggcatcaacaataatttggccactatccagaagtgttgg | 1705 |
| Db | 1261 | tatgtctgatatcctctctggcatcaacaataatttggccactatccagaagtgttgg | 1320 |
| QY | 1706 | cccgcatctgccaanaagcttgaccctctgataaacactglttggagaaatgccaacgtgttc | 1765 |
| Db | 1321 | cccgcatctgccaanaagcttgaccctctgataaacactglttggagaaatgccaacgtgttc | 1380 |
| QY | 1766 | tatatgtcgtcgtcattaaagtlttttgggtgccatttctttacactatcgccaaaagt | 1825 |
| Db | 1381 | tatatgtcgtcgtcattaaagtlttttgggtgccatttctttacactatcgccaaaagt | 1440 |
| QY | 1826 | gaagatgacaaaactggctcctaagatgacacatgtagcaagaagacactga | 1873 |
| Db | 1441 | gaagatgacaaaactggctcctaagatgacacatgtagcaagaagacactga | 1488 |
| RESULT 10 | | | |
| AAZ50880 | | | |
| ID | AAZ50880 | standard; cDNA; 1485 BP. | |
| XX | AAZ50880; | | |
| XX | | | |
| DT | 31-MAY-2000 | (first entry) | |
| XX | | | |
| DE | Human/sheep consensus GBS toxin receptor cDNA sequence-1. | | |
| XX | | | |
| KM | Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59; | | |
| KM | pathological vascularisation; cancer metastases; angiogenesis; | | |
| KM | neovascularisation; reperfusion injury; scarring; keloid; | | |
| KM | chronic inflammation; rheumatoid arthritis; psoriasis; neural injury; | | |
| KM | endothelial cell proliferation; antibacterial; anticancer; | | |
| KM | anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss. | | |
| XX | | | |
| OS | Homo sapiens. | | |
| OS | | | |
| XX | Ovis sp. | | |
| XX | | | |
| FH | Key | Location/Qualifiers | |
| FT | mat_peptide | 1..1485 | |
| FT | | /*tag= a | |
| FT | | /product= "Human/sheep consensus GBS toxin receptor | |
| FT | | protein sequence" | |
| FT | | /transl_except= (pos:4..6; aa:Xaa) | |
| FT | | /transl_except= (pos:7..9; aa:Xaa) | |
| FT | | /transl_except= (pos:16..18; aa:Xaa) | |
| FT | | /transl_except= (pos:22..24; aa:Xaa) | |
| FT | | /transl_except= (pos:28..30; aa:Xaa) | |
| FT | | /transl_except= (pos:31..33; aa:Xaa) | |
| FT | | /transl_except= (pos:34..36; aa:Xaa) | |
| FT | | /transl_except= (pos:46..48; aa:Xaa) | |
| FT | | /transl_except= (pos:49..51; aa:Xaa) | |
| FT | | /transl_except= (pos:58..60; aa:Xaa) | |
| FT | | /transl_except= (pos:61..63; aa:Xaa) | |
| FT | | /transl_except= (pos:64..66; aa:Xaa) | |
| FT | | /transl_except= (pos:67..69; aa:Xaa) | |
| FT | | /transl_except= (pos:70..72; aa:Xaa) | |
| FT | | /transl_except= (pos:73..75; aa:Xaa) | |

FT /transl_except- (pos:79..81; aa:Xaa)
FT /transl_except- (pos:85..87; aa:Xaa)
FT /transl_except- (pos:91..93; aa:Xaa)
FT /transl_except- (pos:100..102; aa:Xaa)
FT /transl_except- (pos:124..126; aa:Xaa)
FT /transl_except- (pos:130..132; aa:Xaa)
FT /transl_except- (pos:136..138; aa:Xaa)
FT /transl_except- (pos:154..156; aa:Xaa)
FT /transl_except- (pos:157..159; aa:Xaa)
FT /transl_except- (pos:163..165; aa:Xaa)
FT /transl_except- (pos:169..171; aa:Xaa)
FT /transl_except- (pos:181..183; aa:Xaa)
FT /transl_except- (pos:187..189; aa:Xaa)
FT /transl_except- (pos:190..192; aa:Xaa)
FT /transl_except- (pos:196..198; aa:Xaa)
FT /transl_except- (pos:202..204; aa:Xaa)
FT /transl_except- (pos:211..213; aa:Xaa)
FT /transl_except- (pos:220..222; aa:Xaa)
FT /transl_except- (pos:223..225; aa:Xaa)
FT /transl_except- (pos:235..237; aa:Xaa)
FT /transl_except- (pos:241..243; aa:Xaa)
FT /transl_except- (pos:244..246; aa:Xaa)
FT /transl_except- (pos:250..252; aa:Xaa)
FT /transl_except- (pos:277..279; aa:Xaa)
FT /transl_except- (pos:280..282; aa:Xaa)
FT /transl_except- (pos:283..285; aa:Xaa)
FT /transl_except- (pos:295..297; aa:Xaa)
FT /transl_except- (pos:304..306; aa:Xaa)
FT /transl_except- (pos:337..339; aa:Xaa)
FT /transl_except- (pos:340..342; aa:Xaa)
FT /transl_except- (pos:346..348; aa:Xaa)
FT /transl_except- (pos:367..369; aa:Xaa)
FT /transl_except- (pos:394..396; aa:Xaa)
FT /transl_except- (pos:397..399; aa:Xaa)
FT /transl_except- (pos:406..408; aa:Xaa)
FT /transl_except- (pos:409..411; aa:Xaa)
FT /transl_except- (pos:412..414; aa:Xaa)
FT /transl_except- (pos:421..423; aa:Xaa)
FT /transl_except- (pos:430..432; aa:Xaa)
FT /transl_except- (pos:433..435; aa:Xaa)
FT /transl_except- (pos:436..438; aa:Xaa)
FT /transl_except- (pos:442..444; aa:Xaa)
FT /transl_except- (pos:445..447; aa:Xaa)
FT /transl_except- (pos:463..465; aa:Xaa)
FT /transl_except- (pos:475..477; aa:Xaa)
FT /transl_except- (pos:481..483; aa:Xaa)
FT /transl_except- (pos:487..489; aa:Xaa)
FT /transl_except- (pos:490..492; aa:Xaa)
FT /transl_except- (pos:493..495; aa:Xaa)
FT /transl_except- (pos:496..498; aa:Xaa)
FT /transl_except- (pos:502..504; aa:Xaa)
FT /transl_except- (pos:514..516; aa:Xaa)
FT /transl_except- (pos:529..531; aa:Xaa)
FT /transl_except- (pos:535..537; aa:Xaa)
FT /transl_except- (pos:562..564; aa:Xaa)
FT /transl_except- (pos:589..591; aa:Xaa)
FT /transl_except- (pos:595..597; aa:Xaa)
FT /transl_except- (pos:598..600; aa:Xaa)
FT /transl_except- (pos:604..606; aa:Xaa)
FT /transl_except- (pos:619..621; aa:Xaa)
FT /transl_except- (pos:634..636; aa:Xaa)
FT /transl_except- (pos:655..657; aa:Xaa)
FT /transl_except- (pos:691..693; aa:Xaa)
FT /transl_except- (pos:694..696; aa:Xaa)
FT /transl_except- (pos:700..702; aa:Xaa)
FT /transl_except- (pos:703..705; aa:Xaa)
FT /transl_except- (pos:706..708; aa:Xaa)
FT /transl_except- (pos:712..714; aa:Xaa)
FT /transl_except- (pos:715..717; aa:Xaa)
FT /transl_except- (pos:724..726; aa:Xaa)
FT /transl_except- (pos:727..729; aa:Xaa)
FT /transl_except- (pos:736..738; aa:Xaa)
FT /transl_except- (pos:748..750; aa:Xaa)

FT /transl_except- (pos:757..759; aa:Xaa)
FT /transl_except- (pos:760..762; aa:Xaa)
FT /transl_except- (pos:769..771; aa:Xaa)
FT /transl_except- (pos:772..774; aa:Xaa)
FT /transl_except- (pos:775..777; aa:Xaa)
FT /transl_except- (pos:778..780; aa:Xaa)
FT /transl_except- (pos:790..792; aa:Xaa)
FT /transl_except- (pos:793..795; aa:Xaa)
FT /transl_except- (pos:811..813; aa:Xaa)
FT /transl_except- (pos:820..822; aa:Xaa)
FT /transl_except- (pos:847..849; aa:Xaa)
FT /transl_except- (pos:850..852; aa:Xaa)
FT /transl_except- (pos:853..855; aa:Xaa)
FT /transl_except- (pos:856..858; aa:Xaa)
FT /transl_except- (pos:862..864; aa:Xaa)
FT /transl_except- (pos:880..882; aa:Xaa)
FT /transl_except- (pos:883..885; aa:Xaa)
FT /transl_except- (pos:892..894; aa:Xaa)
FT /transl_except- (pos:922..924; aa:Xaa)
FT /transl_except- (pos:928..930; aa:Xaa)
FT /transl_except- (pos:943..945; aa:Xaa)
FT /transl_except- (pos:952..954; aa:Xaa)
FT /transl_except- (pos:955..957; aa:Xaa)
FT /transl_except- (pos:970..972; aa:Xaa)
FT /transl_except- (pos:994..996; aa:Xaa)
FT /transl_except- (pos:997..999; aa:Xaa)
FT /transl_except- (pos:1009..1011; aa:Xaa)
FT /transl_except- (pos:1012..1014; aa:Xaa)
FT /transl_except- (pos:1033..1035; aa:Xaa)
FT /transl_except- (pos:1063..1065; aa:Xaa)
FT /transl_except- (pos:1081..1083; aa:Xaa)
FT /transl_except- (pos:1084..1086; aa:Xaa)
FT /transl_except- (pos:1090..1092; aa:Xaa)
FT /transl_except- (pos:1096..1098; aa:Xaa)
FT /transl_except- (pos:1111..1113; aa:Xaa)
FT /transl_except- (pos:1126..1128; aa:Xaa)
FT /transl_except- (pos:1129..1131; aa:Xaa)
FT /transl_except- (pos:1138..1140; aa:Xaa)
FT /transl_except- (pos:1141..1143; aa:Xaa)
FT /transl_except- (pos:1144..1146; aa:Xaa)
FT /transl_except- (pos:1147..1149; aa:Xaa)
FT /transl_except- (pos:1150..1152; aa:Xaa)
FT /transl_except- (pos:1153..1155; aa:Xaa)
FT /transl_except- (pos:1168..1170; aa:Xaa)
FT /transl_except- (pos:1174..1176; aa:Xaa)
FT /transl_except- (pos:1180..1182; aa:Xaa)
FT /transl_except- (pos:1189..1191; aa:Xaa)
FT /transl_except- (pos:1201..1203; aa:Xaa)
FT /transl_except- (pos:1246..1248; aa:Xaa)
FT /transl_except- (pos:1270..1272; aa:Xaa)
FT /transl_except- (pos:1291..1293; aa:Xaa)
FT /transl_except- (pos:1306..1308; aa:Xaa)
FT /transl_except- (pos:1315..1317; aa:Xaa)
FT /transl_except- (pos:1324..1326; aa:Xaa)
FT /transl_except- (pos:1330..1332; aa:Xaa)
FT /transl_except- (pos:1333..1335; aa:Xaa)
FT /transl_except- (pos:1339..1341; aa:Xaa)
FT /transl_except- (pos:1348..1350; aa:Xaa)
FT /transl_except- (pos:1357..1359; aa:Xaa)
FT /transl_except- (pos:1372..1374; aa:Xaa)
FT /transl_except- (pos:1375..1377; aa:Xaa)
FT /transl_except- (pos:1381..1383; aa:Xaa)
FT /transl_except- (pos:1384..1386; aa:Xaa)
FT /transl_except- (pos:1396..1398; aa:Xaa)
FT /transl_except- (pos:1402..1404; aa:Xaa)
FT /transl_except- (pos:1420..1422; aa:Xaa)
FT /transl_except- (pos:1444..1446; aa:Xaa)
FT /transl_except- (pos:1456..1458; aa:Xaa)
FT /transl_except- (pos:1459..1461; aa:Xaa)
FT /transl_except- (pos:1462..1464; aa:Xaa)
FT /transl_except- (pos:1471..1473; aa:Xaa)
FT /transl_except- (pos:1483..1485; aa:Xaa)
FT /note="Xaa = unknown"

KW neurological disorder; ss.
OS Homo sapiens.
XX MO200153455-A2.
XX 26-JUL-2001.
XX 22-DEC-2000; 2000MO-US35017.
XX 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX (HYSE-) HYSEQ INC.
PI Tang YT, Liu C, Drmanac RT;
XX WPI: 2001-457603/49.
DR P-PSDB; AAM25685.
PT Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX Claim 1; Page 537; 1217pp; English.
XX AAH9166 to AAH9904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virocidic; anti-HIV; fungicidal; antimutagen;
CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnery;
CC antulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytosolic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX Sequence 1975 BP; 535 A; 437 C; 434 G; 569 T; 0 other:
SQ

Query Match 48.0%; Score 1405.6; DB 22; Length 1975;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1423; Conservative 0; Mismatches 9; Indels 2; Gaps 1;

QY 1479 gaattttagccttataaggaatgatgagcctgcagatcttcctgtagctgctgctca 1538
DB 1975 GAATTTTAGCCTTATAGGAATGATGAGACCTGCAGATATTCCTGAGTGGCTGCTTCA 1916
QY 1539 ttgagctgtgattatctcttgccgctgtgttccttaactatatacaacactggagagct 1598
DB 1915 TTGAGCTGTGATTATCTTGGCCGCTTGTCTTAACATATACAAACACCTGGAGGCT 1856
QY 1599 ttgctctcttgatttagcatcaacacatctgatatcttcctctgtatgctgatatcc 1658
DB 1855 TTGCTCTCTTGATTAGCATCAACCATCTGATATTCTCTCTGTAAGCTGGATATCC 1796
QY 1659 tccctggacacacaataacattgccaactatccagaatggttgagccgctcaattgcta 1718
DB 1795 TCCTGGGACATCAACAATACATTGGCACATATCCAGAGATGGTGGCCCGCATTTGCTA 1736
QY 1719 aaagtcgacccctgataacactgttgagaaatggcaaacgtytctatattgctgctg 1778
|||||

DB 1735 AAGCTGACCCCTGATTAACACTGTTGGAGAAATGGCAAAACCGTGTCTATATTGCTGCTG 1676
QY ctatiaatgtttttgtgctcatttctta--cactatccgcaaaagtgtacataaa 1836
DB 1675 CTATTAATGTTTTTGGTGGCATTTTCTTAACACTTTTCGCCAAAGGTGAAGTACAAAA 1616
QY ctggagctcctaaatgaatcaacatgagacacagacactgaaggaacaaataaatactcctgc 1896
DB 1615 CTGGGCTCAATGATGACCATGACACAGACACTGAAGAACCAATTAATATCTCGCC 1556
QY tctattaaatgatttattatcatatgataccctcaagtgccctctgatttggtaagca 1956
DB 1555 TCTATTAAATGATTTTTTATTTATCATGATGATCCCAAGTCCCTGCTATTTGGTAAGCA 1496
QY ttctatgctcttttttaattgtaactgtatagattttaaagccataatcaatgaata 2016
DB 1495 TTCTATGCTCTTTTAAATGTAAGTGTATGATTTTAAAGCCATATATCATGAAATA 1436
QY tcaatagttgccaagaataaataaataaactgtgtttaaataatgaataatgaagctag 2076
DB 1435 TCACATGTTGGCAGATTAATAAATGAACGTGTGTTAATTATGAATAATATGTAACCTAG 1376
QY gaattctacttagtgatcaatcactcgtcgtcgtcagtcgaggaacatgaaatggaagat 2136
DB 1375 GACTTCTACTTTAGGTTCAACATACCTGCTGTAGTGGGCAACATGAAGTAGACAGTT 1316
QY ctgtgattttttttagggccaataaaggaatgaagctgaagaacccctccgatacctt 2196
DB 1315 CTGTATATTTTAAAGGCGCATTAAGAAGGATGAGCTGAACAGCTCTGATATCCTT 1256
QY tgcctaaataactatagataaattctcagagtaactgataaacaactgtgttgttcaact 2256
DB 1255 TGCTTATTAATCACTAGATGATATATTCAGTAGTACTGATTAACACCTGTTGTTGACTT 1196
QY tctctataaataatgtcagctctctctgacacttagaactcaacttagactctctg 2316
DB 1195 TCCTCTATAAATTTGTCACCTCTCTGTGACCTTTAAGCTCAACCTTGTGACTTCTGTG 1136
QY gaggctcactcaactgtataatttcgctggcaacgtgagactgaaggaatgtgcccagagc 2376
DB 1135 GACCTGCCATCCACTTATATTTGCGCTGGCAACGTGAGCTGAGGAGATGTGCCAGGC 1076
QY agcttgcgaagcactccctccctgctgctcagagtcagagtgcccagtcgttatcaagaagca 2436
DB 1075 AGCTGCCAAGCACTCCCTCCCTGGCTTCAAGGCGCAAGTAGGCCAGCGTTTACAGAGCA 1016
QY gcatcgaagcccaagagcaggtgtgagactcttgagctgagcttccctcttgaggggctat 2496
DB 1015 GCATCCAAAGCCAGACCAAGTGCACCTCTTGGCTGTGCTTCTCTGAGGGGCTAT 956
QY caatgtgtatataaagcccttgagtaagcaagaacgtgagatccactgctatgctctga 2556
DB 955 CAATGTGTATGATTAAGCCCTGAGTAGCAAGACATGAGATGCCATGCTATAGTCTTTGA 896
QY tcaatcctcaaaccttccctccccaagacagaagaaatattgctgcatgcaactgcaaa 2616
DB 895 TACATCTCAAACTTCCCTCCACACAGAGAAATATTGCTGGCATGCAACCTGCAGAA 836
QY aagaataaataatgcaagc 2676
DB 835 AAGAAAAATGCAAGAGGGCGGGGCAAGGTGCTCAAGCTGTAAATCCAGCACTTGGGG 776
QY ggcctaggtgtggcgaataatgagatcaagagatctgagaccagcctggccaagatggtga 2736
DB 775 GGTGAGGTGGCGAATTCATGAGATCAGAGATTCCAGACCACTGGCCAGCATGGTGA 716
QY accccatcctactaataaataacaaaatlaagctggcgctgtgcaagcgccgctgaatc 2796
DB 715 ACCCATCTCTCTATAAATTAACAAAAATTTACTGGGCGCTGTGACGGCGCTGTAAATC 656
QY ccaagataactcaggaagctgagagtgaggaagaaatcaacttgaactcggagagtggaagtgcag 2856
DB 655 CCAGATACTCAGAGAGCTGAGGTAGAGATCACTTGAACCTTGGAGAGATGAAGTTTGCAG 596
|||||

OY 2857 tgaccagatcacgccactgcaccagctggcgatggagcgagactccaa 2910
|||||
Db 595 tgaccagatcacgccactgcaccctggcgatggagcgagatcccca 542

RESULT 12
AA250881
ID AA250881 standard: cDNA; 1485 BP.
XX
AC AA250881;
XX
DI 31-MAY-2000 (first entry)
XX
DE Human/sheep consensus GBS toxin receptor cDNA sequence-2.
XX
KW Human GBS toxin receptor; group B beta-hemolytic streptococci; Hp59;
KW pathological vascularisation; cancer metastases; angiogenesis;
KW neovascularisation; reperfusion injury; scarring; keloid;
KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW endothelial cell proliferation; antibacterial; anticancer;
KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.
XX
OS Homo sapiens.
XX
Ovis sp.
FH
FH Key mat_peptide
FT 1.1485
FT Location/Qualifiers
FT /tag= a
FT /product= "Human/sheep consensus GBS toxin receptor
FT protein sequence"
FT /transl_except- (pos:4..6; aa:Xaa)
FT /transl_except- (pos:7..9; aa:Xaa)
FT /transl_except- (pos:16..18; aa:Xaa)
FT /transl_except- (pos:22..24; aa:Xaa)
FT /transl_except- (pos:28..30; aa:Xaa)
FT /transl_except- (pos:31..33; aa:Xaa)
FT /transl_except- (pos:34..36; aa:Xaa)
FT /transl_except- (pos:46..48; aa:Xaa)
FT /transl_except- (pos:49..51; aa:Xaa)
FT /transl_except- (pos:58..60; aa:Xaa)
FT /transl_except- (pos:61..63; aa:Xaa)
FT /transl_except- (pos:64..66; aa:Xaa)
FT /transl_except- (pos:67..69; aa:Xaa)
FT /transl_except- (pos:70..72; aa:Xaa)
FT /transl_except- (pos:73..75; aa:Xaa)
FT /transl_except- (pos:79..81; aa:Xaa)
FT /transl_except- (pos:85..87; aa:Xaa)
FT /transl_except- (pos:91..93; aa:Xaa)
FT /transl_except- (pos:100..102; aa:Xaa)
FT /transl_except- (pos:124..126; aa:Xaa)
FT /transl_except- (pos:130..132; aa:Xaa)
FT /transl_except- (pos:136..138; aa:Xaa)
FT /transl_except- (pos:154..156; aa:Xaa)
FT /transl_except- (pos:157..159; aa:Xaa)
FT /transl_except- (pos:163..165; aa:Xaa)
FT /transl_except- (pos:169..171; aa:Xaa)
FT /transl_except- (pos:181..183; aa:Xaa)
FT /transl_except- (pos:187..189; aa:Xaa)
FT /transl_except- (pos:190..192; aa:Xaa)
FT /transl_except- (pos:196..198; aa:Xaa)
FT /transl_except- (pos:202..204; aa:Xaa)
FT /transl_except- (pos:211..213; aa:Xaa)
FT /transl_except- (pos:220..222; aa:Xaa)
FT /transl_except- (pos:223..225; aa:Xaa)
FT /transl_except- (pos:235..237; aa:Xaa)
FT /transl_except- (pos:241..243; aa:Xaa)
FT /transl_except- (pos:244..246; aa:Xaa)
FT /transl_except- (pos:250..252; aa:Xaa)
FT /transl_except- (pos:277..279; aa:Xaa)
FT /transl_except- (pos:280..282; aa:Xaa)
FT /transl_except- (pos:283..285; aa:Xaa)
FT /transl_except- (pos:295..297; aa:Xaa)

FT /transl_except- (pos:304..306; aa:Xaa)
FT /transl_except- (pos:337..339; aa:Xaa)
FT /transl_except- (pos:340..342; aa:Xaa)
FT /transl_except- (pos:346..348; aa:Xaa)
FT /transl_except- (pos:367..369; aa:Xaa)
FT /transl_except- (pos:394..396; aa:Xaa)
FT /transl_except- (pos:397..399; aa:Xaa)
FT /transl_except- (pos:406..408; aa:Xaa)
FT /transl_except- (pos:409..411; aa:Xaa)
FT /transl_except- (pos:412..414; aa:Xaa)
FT /transl_except- (pos:421..423; aa:Xaa)
FT /transl_except- (pos:430..432; aa:Xaa)
FT /transl_except- (pos:433..435; aa:Xaa)
FT /transl_except- (pos:436..438; aa:Xaa)
FT /transl_except- (pos:442..444; aa:Xaa)
FT /transl_except- (pos:445..447; aa:Xaa)
FT /transl_except- (pos:463..465; aa:Xaa)
FT /transl_except- (pos:475..477; aa:Xaa)
FT /transl_except- (pos:481..483; aa:Xaa)
FT /transl_except- (pos:487..489; aa:Xaa)
FT /transl_except- (pos:490..492; aa:Xaa)
FT /transl_except- (pos:493..495; aa:Xaa)
FT /transl_except- (pos:496..498; aa:Xaa)
FT /transl_except- (pos:502..504; aa:Xaa)
FT /transl_except- (pos:514..516; aa:Xaa)
FT /transl_except- (pos:529..531; aa:Xaa)
FT /transl_except- (pos:535..537; aa:Xaa)
FT /transl_except- (pos:562..564; aa:Xaa)
FT /transl_except- (pos:589..591; aa:Xaa)
FT /transl_except- (pos:595..597; aa:Xaa)
FT /transl_except- (pos:598..600; aa:Xaa)
FT /transl_except- (pos:604..606; aa:Xaa)
FT /transl_except- (pos:619..621; aa:Xaa)
FT /transl_except- (pos:634..636; aa:Xaa)
FT /transl_except- (pos:653..657; aa:Xaa)
FT /transl_except- (pos:691..693; aa:Xaa)
FT /transl_except- (pos:694..696; aa:Xaa)
FT /transl_except- (pos:700..702; aa:Xaa)
FT /transl_except- (pos:703..705; aa:Xaa)
FT /transl_except- (pos:706..708; aa:Xaa)
FT /transl_except- (pos:712..714; aa:Xaa)
FT /transl_except- (pos:715..717; aa:Xaa)
FT /transl_except- (pos:724..726; aa:Xaa)
FT /transl_except- (pos:727..729; aa:Xaa)
FT /transl_except- (pos:736..738; aa:Xaa)
FT /transl_except- (pos:748..750; aa:Xaa)
FT /transl_except- (pos:757..759; aa:Xaa)
FT /transl_except- (pos:760..762; aa:Xaa)
FT /transl_except- (pos:769..771; aa:Xaa)
FT /transl_except- (pos:772..774; aa:Xaa)
FT /transl_except- (pos:775..777; aa:Xaa)
FT /transl_except- (pos:778..780; aa:Xaa)
FT /transl_except- (pos:790..792; aa:Xaa)
FT /transl_except- (pos:793..795; aa:Xaa)
FT /transl_except- (pos:811..813; aa:Xaa)
FT /transl_except- (pos:820..822; aa:Xaa)
FT /transl_except- (pos:847..849; aa:Xaa)
FT /transl_except- (pos:850..852; aa:Xaa)
FT /transl_except- (pos:853..855; aa:Xaa)
FT /transl_except- (pos:856..858; aa:Xaa)
FT /transl_except- (pos:862..864; aa:Xaa)
FT /transl_except- (pos:880..882; aa:Xaa)
FT /transl_except- (pos:883..885; aa:Xaa)
FT /transl_except- (pos:892..894; aa:Xaa)
FT /transl_except- (pos:922..924; aa:Xaa)
FT /transl_except- (pos:928..930; aa:Xaa)
FT /transl_except- (pos:943..945; aa:Xaa)
FT /transl_except- (pos:952..954; aa:Xaa)
FT /transl_except- (pos:955..957; aa:Xaa)
FT /transl_except- (pos:970..972; aa:Xaa)
FT /transl_except- (pos:994..996; aa:Xaa)
FT /transl_except- (pos:997..999; aa:Xaa)
FT /transl_except- (pos:1009..1011; aa:Xaa)

| Query Match | 44.5% | Score 1305; | DB 21; | Length 1485; |
|---|-------|---------------------|-----------|--------------|
| Best Local Similarity | 87.9% | Pred. No. 1.1e-310; | | |
| Matches 1305; Conservative | 0; | Mismatches 180; | Indels 0; | Gaps 0; |
| 386 atgaagctcgcggtccgcgaacctgcgcgaagatggcgaagagacgaaccgcaacy | 445 | | | |
| 1 atgaagtcnccggttcnmgacttngccnngancgngcgaaggaagngcncmgaccgcaacn | 60 | | | |
| 446 cctctctacacggcgcccccacggcgcaagcgctcagltgtgtctgtctgtcgttac | 505 | | | |

[illegible]

| | | | |
|-----------|---|--|------|
| Oy | 1060 | ttgactatgctctcactcttcttgtagtactatggaatcttggctcttcttgat | 1119 |
| Db | 793 | ttgactaatgctctcactcttcttgtagtactatggaatcttggctcttcttgat | 852 |
| Oy | 1120 | ctggttgaattgtagcaccacacaaaacacagaagatctcccatatgaaagata | 1179 |
| Db | 853 | ctggttgaattgtagcaccacacacaaaacacagaagatctcccatatgaaagata | 912 |
| Oy | 1180 | catctcttcacatctaa | 1196 |
| Db | 913 | catctcttcacatctaa | 929 |
| RESULT 14 | | | |
| AAK93901 | | | |
| ID | AAK93901 | standard; cDNA; 853 BP. | |
| AC | AAK93901; | | |
| XX | | | |
| DT | 06-NOV-2001 | (first entry) | |
| XX | | | |
| DE | Human cDNA clone representative sequence, SEQ ID NO: 2361. | | |
| XX | | | |
| XX | Human; full length cDNA; cDNA synthesis; oligo-capping; ss. | | |
| OS | Homo sapiens. | | |
| XX | | | |
| PN | EP1130094-A2. | | |
| PD | 05-SEP-2001. | | |
| XX | | | |
| PF | 07-JUL-2000; 2000EP-0114089. | | |
| XX | | | |
| PR | 08-JUL-1999; 99JP-0194486. | | |
| PR | 11-JAN-2000; 2000JP-0118774. | | |
| PR | 02-MAY-2000; 2000JP-0183765. | | |
| PA | (HELI-) HELIX RES INST. | | |
| XX | | | |
| PI | Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y; | | |
| PI | Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H; | | |
| XX | | | |
| DR | WPI; 2001-524255/58. | | |
| PT | 830 Primers useful for synthesizing full length cDNA clones and their | | |
| XX | use in genetic manipulation - | | |
| XX | | | |
| XX | Example 11; SEQ ID NO 2361; 1360pp + sequence listing; English. | | |
| CC | The invention relates to primers for synthesizing full length cDNA | | |
| CC | clones. 830 cDNA molecules encoding a human protein have been | | |
| CC | isolated and nucleotide sequences of 5' and 3'-ends of the cDNA | | |
| CC | molecules have been determined. Primers for synthesizing the full length | | |
| CC | cDNA are useful for clarifying the function of the protein encoded by | | |
| CC | the cDNA. The full length clones were obtained by construction of full | | |
| CC | length enriched cDNA libraries that were synthesised by the oligo-capping | | |
| CC | method. The primers enable the production of the full length cDNA easily | | |
| CC | without any special methods. The present sequence was used as the | | |
| CC | representative sequence from a human clone which was used in | | |
| CC | homology searches to identify the clone. | | |
| CC | Note: The sequence data for this patent did not form part of the printed | | |
| CC | specification, but was obtained in CD-ROM format directly from EPO. | | |
| XX | | | |
| XX | Sequence 853 BP; 198 A; 200 C; 213 G; 239 T; 3 other; | | |

| | | | | |
|-----------------------|-----------------|---------------------|-----------|-------------|
| Query Match | 28.3%; | Score 827.8; | DB 22; | Length 853; |
| Best Local Similarity | 98.7%; | Pred. No. 1.7e-193; | | |
| Matches 843; | Conservative 0; | Mismatches 10; | Indels 1; | Gaps 1; |

| | | | |
|----|-----|---|-----|
| QY | 304 | agctcgcgccggggttcgaggctgtccgcgcgcctccctctctcgcaagtgcgag | 363 |
| Db | 1 | agctcgcgccggggttcgaggctgtccgcgcgcctccctctctcgcaagtgcgag | 60 |

| | | | |
|----|------|---|------|
| QY | 364 | tacacctgcttaagtaagcggtcgaagaggtcttcgcggtttcgagaaactctgccccgaaagatg | 423 |
| Db | 61 | tacacctcttcaagtaagtcggtcgaagaggtcttcgcggtttcgagaaactctgccccgaaagatg | 120 |
| QY | 424 | cgaagagagacacggaacgcgaacgcctctcttcacaggcgccccacgagccggaagcgcgtcc | 483 |
| Db | 121 | cgaagagagacacggaacgcgaacgcctctcttcacaggcgccccacgagccggaagcgcgtcc | 180 |
| QY | 484 | agtgctgcgtcctgcgtcgttaacaactatgacaatttggccctttttgttttcttcaatgt | 543 |
| Db | 181 | agtgctgcgtcctgcgtcgttaacaactatgacaatttggccctttttgttttcttcaatgt | 240 |
| QY | 544 | gtatccatctagctgtgaatctcgaagtgctgttgatggatgatgatgatgattcaataaac | 603 |
| Db | 241 | gtatccatctagctgtgaatctcgaagtgctgttgatggatgatgatgatgattcaataaac | 300 |
| QY | 604 | tttagaagatataatgaactctccaagcggtgttccagaagcatctctgctcccatataaagtca | 663 |
| Db | 301 | tttagaagatataatgaactctccaagcggtgttccagaagcatctctgctcccatataaagtca | 360 |
| QY | 664 | tcataatcaaaacgggttaagaagtaaccaatggtgatgcaagaactccaagatgtagttctcgg | 723 |
| Db | 361 | tcataatcaaaacgggttaagaagtaaccaatggtgatgcaagaactccaagatgtagttctcgg | 420 |
| QY | 724 | tttccttttttatgtgcacacatcaacaacagatctctctggagatgatgtgttcgacgaatat | 783 |
| Db | 421 | tttccttttttatgtgcacacatcaacaacagatctctctggagatgatgtgttcgacgaatat | 480 |
| QY | 784 | agggggggaaaaatgtctgtctagaagatttggagatcctcttggcactgctgtcctcaaccgttcaac | 843 |
| Db | 481 | agggggggaaaaatgtctgtctagaagatttggagatcctcttggcactgctgtcctcaaccgttcaac | 540 |
| QY | 844 | tcccatgtgctgcgaatttaaggagtttggacaactcatcttgaacctcgaagagacatgaaggact | 903 |
| Db | 541 | tcccatgtgctgcgaatttaaggagtttggacaactcatcttgaacctcgaagagacatgaaggact | 600 |
| QY | 904 | aggagagaggtgttacaatttccacgaacatgacatgccaatgtgctctctcttggagcccccct | 963 |
| Db | 601 | aggagagaggtgttacaatttccacgaacatgacatgccaatgtgctctctcttggagcccccct | 660 |
| QY | 964 | tgaagaagacaaactctcttgatcattcgtatgcaagagacacagctcttggagacgataatc | 1023 |
| Db | 661 | tgaagaagacaaactctcttgatcattcgtatgcaagagacacagctcttggagacgataatc | 720 |
| QY | 1024 | tcttcctcttctcgaaataatttgctacataatgcaattggaactatgactatgctcttactctt | 1083 |
| Db | 721 | tcttcctcttctcgaaataatttgctacataatgcaattggaactatgactatgctcttactctt | 780 |
| QY | 1084 | tgtgactattggaatattttgtttctcttgtgtgacatcgtttagttagtgaacacacca | 1143 |
| Db | 781 | tgtgactattggaatattttgtttctcttgtgtgacatcgtttagttagtgaacacacca | 839 |
| QY | 1144 | aaaacacaaagagaa 1157 | |
| Db | 840 | aaaacacaaagagaa 853 | |

| | |
|----------|----------------------------------|
| RESULT | 15 |
| AAK92364 | |
| ID | AAK92364 standard; cDNA; 838 BP. |

06-NOV-2001 (first entry)
Human CDNA 5'-end sequence, SEQ ID NO: 824.

KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

| | |
|----|---------------|
| AA | Homo sapiens. |
| OS | |
| XX | |
| PN | EP1130094-A2. |

...

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 15, 2002, 09:40:35 ; Search time 6938.09 Seconds
(without alignments)
8837.411 Million cell updates/sec

Title: US-09-776-865-1

Perfect score: 2930

Sequence: 1 gtcggtcgaagccctcccc.....ctcaaaaaaaaaaaaaaaaaa 2930

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
|------------|-------|-------------|--------|-------|-------------|

ALIGNMENTS

| RESULT | 1 | AX207624 | Sequence 1 from Patent WO0156598. | 2930 bp | DNA | linear | PAT 31-AUG-2001 |
|----------|------------|--|-----------------------------------|---------|-----|--------|-----------------|
| AX207624 | LOCUS | AX207624 | Sequence 1 from Patent WO0156598. | | | | |
| | DEFINITION | AX207624 | Sequence 1 from Patent WO0156598. | | | | |
| | ACCESSION | AX207624 | Sequence 1 from Patent WO0156598. | | | | |
| | VERSION | AX207624.1 | GI:15422329 | | | | |
| | KEYWORDS | human. | | | | | |
| | SOURCE | human. | | | | | |
| | ORGANISM | Homo sapiens | | | | | |
| | REFERENCE | 1 (bases 1 to 2930) | | | | | |
| | AUTHORS | Hellergqvist, C.G. | | | | | |
| | TITLE | Methods for preventing or attenuating pathoangitogenic conditions by using the gbs-toxin (cm101) receptor as a vaccine | | | | | |
| | JOURNAL | Patent: WO 0156598-A 1 09-AUG-2001; | | | | | |
| | FEATURES | VANDERBILT UNIVERSITY (US) | | | | | |
| | Source | Location/Qualifiers | | | | | |
| | CDS | 1..2930 | | | | | |
| | | /organism="Homo sapiens" | | | | | |
| | | /db_xref="taxon:9606" | | | | | |
| | | 263..1873 | | | | | |
| | | /note="unnamed protein product" | | | | | |

```
/codon_start=1
/protein_id="CAC60358.1"
/db_xref="GI:15422330"
/translation="MAAGAMPVPPVQPARPGFGLSGRRSLILCOVASTPAHVGVNRS
PVLDLRNDEESTDRTPULPGAPRAEAPVCCSARYNLIAIAFFGFFIVYALRVNLS
VALVMDVNDSTEDRTKPLKAPESAPIKVHHNQTGRKYOWDAETQGWILGSFFYGY
ITIQIPGVASKIGKMLIGLITGLTAVLFTPIAADLGVGPIVLRLALGELGEV
TFPAMHAWSSWAPLERSKLLISVYGAOLGTVISLPLSGIICYVMNWTYVYFPGT
IGFWFLWILVSDTPPOKHRIKSHYEXEYILSSRLNLSOKSPVWPVILASLPLWA
IVVAHESYNWTFYLLTLPTYMKEILRNVOENGFLSLPYLGLSWLMLSGQAADN
LRKAMNCTVRRIFSLGIMGPVFLVAAGFIGDISLAVAFITISTILGFCSSG
FSINHLIDPLFAGVILGINTNFATIPGMVGPVIAKSLTPDNTVGEWQTFVIAAAN
VFAGIFDTLFAKGEVONWALNDHGHHRH"
BASE COUNT      729 a   698 c   681 g   822 t
ORIGIN
```

Query Match 100.0%; Score 2930; DB 6; Length 2930;

Best local similarity 100.0%; Pred. No. 0;

Matches 2930; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 qtcggtcgaagccctcccttaattatgtgaattcaagtcceactgcccgccgcaa 60
Db 1 GTTCGGTGAAGCCCTCCCTTAATATGTGCAATTCAAGTCCCCACGTGCCCGCCGCAA 60
Qy 61 gcccactcatctcgtcgtcgccgaggggtggccctgcaactttacaagggggtcgagga 120
Db 61 GCCCCCACTCATCTCGTCTCGGCGGAGGTGGCCCTGCACATTACAAGGGGTGCAGGA 120
Qy 121 gcggagacggtcgtccgaacacggctcccccggcatggtagaccggggcgagcg 180
Db 121 GCGGGAGACGCTGTCGGAACACGGCTCCCGGCATGCGTAGACCGGGCGGAGCGG 180
Qy 181 gctcaattgcgcaatctacagaaactcccaagaactccggttccctagtcacaaaccaa 240
Db 181 GCTACATTTTCGGCCATCTTACGAGAACTCCCGAAGTTCGGCTTCCCTAGTCCAAACCCAA 240
Qy 241 gccagagtgccacacactaaagtgcggcgggcgagtagaaccccccgcgggtcca 300
Db 241 GCCAGAGTTGCCACACCTTAAGATGCGGGCGGGGCGATGACACCGCCCGCCGCTCCA 300
Qy 301 gccagtcggccgggggttcgggtgtcgggcgccggtccctctctcgcagtgcc 360
Db 301 GCCAGCTCGGCCGGGGGTTCGGGGTGTGGGGCGGGCGCTCTCTCTGCGAGGTGGC 360
Qy 361 gactacactgctcagctagggctcatgaggtctccggttcagacctgcccgcgaacga 420
Db 361 GAGTACACCTGCTCAGTAGGCGTCAATGAGGTCTCCGGTTCGAGACTGGCCCGGAACGA 420
Qy 421 tggcaggagacacggaacgcgcctctctacggggcgcccccagggcggaagccgc 480
Db 421 TGGCGAGGAGACGACGACCGCACGCGCTCTTCTACGGGGCGCCACGCGCGCAAGCGCG 480
Qy 481 tcagtgctgctgctgctgttaacacttagcaatttggccctttttgggtttctcat 540
Db 481 TCCAGTGTGCTGCTGCTGCTGTACAACTTAGCAATTTTGGCCCTTTTGGTTTCTTCAT 540
Qy 541 tgtgatgcaattacgtgtaactgaggtgtcggttagtgatagtgatagtgatcaaatc 600
Db 541 TGTGATGCAATACGTGTGTAATCTGAGTGTGCGTTAGTGGATATGATGATTCAAATAC 600
Qy 601 aactttagaagataatagaacttccaagcggtgtccagagcattctgctcccaataaagt 660
Db 601 AACTTTAGAAGATAATAGAACTTTCCAAGCGGTGTCCAGAGCATTTCTGCCATAAAGT 660
Qy 661 tcatcataataacacgggtaagaagtaccatggggtgagaaactcaaggatggattct 720
Db 661 TCATCATATAATCAACCGGTAAAGAAGTACCAATGGGATGAGAAACTCAAGGATGGATTCT 720
Qy 721 cgggtctctttttatggtcatcatcacagattcctcgagagatgtgtgcccagcaa 780
Db 721 CGGTTCTCTTTTATGCGGTACATCATACACAGATTTCTTGGAGGATATGTTGCCAGCAA 780
```

```
Qy 781 aataggggggaaaaactgctgtaggatttgggacatcttggcactgctgtcctaccctgtt 840
Db 781 AATAGGGGGGAAANCTGCTGCTAGGATTTGGGATCTTGGCACTGCTGTCTCACCTGTT 840
Qy 841 cactcccatctgctgcagatttaggaggttggaccactcattgtactcagagcactagaag 900
Db 841 CACTCCCATTTGCTGCAGATTTAGGAGTTGGACCTCATTTGTACTCAGAGCAGCTAGAAG 900
Qy 901 actagagaggggtgttacatttccacgcaatgcatgcatggtgtgttcttgggtcccc 960
Db 901 ACTAGGAGAGGGGTGTACATTTCCACCCATGATGCATGTGGTCTCTTCTTGGGCTCCCC 960
Qy 961 tcttgaaagaagcaaaactctcttagcatttcgtatgcaggagcagcgttgggacagtaat 1020
Db 961 TCTTGAAGAAGCAAACTTCTTAGCATTTTCGTATGAGGAGCACACGTTGGGACACTAAT 1020
Qy 1021 tctctctctctcttcttggaaataattgtctactatatgaattggagacttatgtcttactt 1080
Db 1021 TTCTCTTCTCTTCTTCTTGGAAATATTTGCTACTATATGAATTTGGACTATGTCTTCTACTT 1080
Qy 1081 ttttgcactatttgaataatttgggttcttcttggatctggttagttagtgacacac 1140
Db 1081 TTTTGTACTATTGGAAATATTTGGTTCTTTTGTGGATCTGGTTAGTTAGTAGTACACACC 1140
Qy 1141 acaaaacacagagaaatttcccatatgaaaaaggaaatacattcttctcatcattaagaaa 1200
Db 1141 ACAAAAACACAGAGAAATTTCCCATTTATGAAAAGGAATACATTTCTTTTCATCATTAAGAAA 1200
Qy 1201 tcaagttcttccacagaagtcagtcggtgggtaccacttttbaaaatccctgcacatttg 1260
Db 1201 TCAGCTTTCTTTCACAGAAAGTCAGTGCCTGGGTACCATTTTAAAAATCCCTGCGCACTTTG 1260
Qy 1261 gactatcgtagttgacacatttcttcaactggagactttttatacttttatacaattatt 1320
Db 1261 GGTATCGTAGTTGACACACTTTCTTACAACTGGACATTTTATATACTTTATGACATATT 1320
Qy 1321 gctacttatataagagagatccaaaggttcaagtccaagagaagtggttttttactcttc 1380
Db 1321 GCCTACTTATATGAAGGAGATCCTAAGGTTCAATGTTCAAGAGAATGGGTTTTTATCTTTC 1380
Qy 1381 attgcctatttaggtctgtgttatgtatgatcctcgtctgtgctgaagctgctgacaattt 1440
Db 1381 ATTGCTTTATTTAGGCTCTTGGTTATGTATGATCTCTGTCTGTCAGAGCTGCTGACAAATTT 1440
Qy 1441 aaaggcaaaatgaaattttcaactttatgttccagaaatttttagcctctatagaagat 1500
Db 1441 AAGGGCAAAATGGAATTTTCAACTTTATGTGTGCCAGAAATTTTAGCCCTTATAGGAAT 1500
Qy 1501 gattggacctgcagttatcctcgtgtagctgctggttcattgctggtgattattcttgggc 1560
Db 1501 GATTGGACCTGCAGTATTCTGCTAGTCTGCTGGCTTCATTGGCTGTGATTTATTTCTTTGCG 1560
Qy 1561 cgttcttccatactatacaacacactggaggcttttctctctctctctctctctctctct 1620
Db 1561 CGTTGCTTTTCCTAACATATATCAACACACTGGGAGGCTTTTGTCTCTCTGGAATTTAGCAT 1620
Qy 1621 caaccatctggatattgctcctcgtatcgtatcctcctcctggcgcacacaaatacatt 1680
Db 1621 CAACCATCTGGATATGTCTCTTCTGCTATGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCT 1680
Qy 1681 tggccactattccaggaatggttggcccgctcattgctaaaaagctcgcacccctgataaac 1740
Db 1681 TGCCACTATTTCAGGAATGTTGGGCGCGCTCATTTGTAAGAGTCTGACCCCTGATAAACAC 1740
Qy 1741 tgttgagaatggcaaacccgtgttctataattgctgctgtatttaattgttttgggtccat 1800
Db 1741 TGTTGAGAATGGCAACCCGTGTTCTATATTGCTGCTGTATTAATGTTTGTGGTGCCAT 1800
Qy 1801 tttcttcaactattcgcacaaaggtagaagtaaaaaactgggtcctcaatgatcaccatgg 1860
Db 1801 TTTCTTACACTATTGCGCAAAAGGTGAAGTACAAAACCTGGGCTCTCAATGATCACCATGG 1860
Qy 1861 acacagacactgaaggaaaccaaataaactcctgctctatttaattttttatttattc 1920
```

Db 1861 ACACAGACACTGAAGAACCAATAAATAATCCCTGCTCTATTATGATATTTTATTATATC 1920
QY 1921 atgaacctcaagtcctctgtattgtgtaagcattctatgtctcttttttaattgtac 1980
Db 1921 ATGTAACTCAAGTCCCTCTGTATGTGTAGCAATCTATGCTCTTTTAAATGTATC 1980
QY 1981 ttgtattagatttttaagccctataatcatgaataatcactagttgccagaataataaaa 2040
Db 1981 TTGTATTAGATTTTAAAGCCCTAATATCATGAATAATCACTAGTCCGAGAATAATAAA 2040
QY 2041 tgaactgtgttaattataataatgttaagctaggaactctcacttttaggttcaacatc 2100
Db 2041 TGAACCTGTGTTAAATATGAATGAATGAAGCTAGGACTTCTACTTTAGGTTTCACATC 2100
QY 2101 ctgctgtcgtagtcggcaacatgaatagacagttctcttgaatttttagggccatact 2160
Db 2101 CTGCCCTGTAGTCGGCAGACATGAAGTAGACAGTCTGTGTGATTTTATAGGCCATATC 2160
QY 2161 aaaggaatgagctgaacacagactcctgatacctttgtcttaataaactagatgataat 2220
Db 2161 AAAGGAATGAGCTGAACAGACCTCCTGATACCTTTGCTTAATTAACATAGATGATAAT 2220
QY 2221 tctcaggtactgataaaccctgttgttcttctcacttctcctataaataatgtcagctctc 2280
Db 2221 TCTCAGGTACTGATAAACACCTGTTGTTGTTCACTTTCTCTCATAAAAATTTGTCAGCTCTC 2280
QY 2281 ttgacacttagacctcaacttttagcatctctgtgagctgcatccactgataattt 2340
Db 2281 TCTGACACTTAGACCTCAAACTTTAGCATCTCTGTGGAGTGCATCCACTGATTAATTT 2340
QY 2341 cgctggcaactggaactgagggagtggtgccaggcagctgccaaactcctccctgg 2400
Db 2341 CGCCTGGCAACTGGACTGAGGGGAGTGTGCCAGGCAGCTGCCAAGCACTTCCCTCCCTGG 2400
QY 2401 cttcagggtcagagtgccagcgttttatcagagcagcagcattccaaagccagagccagtgc 2460
Db 2401 CTTTCAGGGTTCAGAGTCCCGACGCTTTATCAGAGGCAGCATCCAAAGCCGACGACGTGC 2460
QY 2461 gactctcgctgtgtccttctctgagggggtatcaatgtgtagataaagccctgagt 2520
Db 2461 GACTCTTCGCTGTGCTCTCTCTGAGGGGTATCAATGCTAGATAAAGCCCTGAGT 2520
QY 2521 aggaagagcagtgagatccactgctatggtttgatatcatctcctcctccctccc 2580
Db 2521 AGGCAAGAGCAGTGAGATCCACTGCTATGCTCTTGATACATCTCAAACTTTCCCTTCCC 2580
QY 2581 agcaagaggaatatgtgctggcatgcaacctgcaaaaaaataatcggaagcgccgggc 2640
Db 2581 AGCACAGAGGAATTTGGTGGCATCAACCTGCAAAAGAAAAAATCGAAGCGCCGGGC 2640
QY 2641 acggtgctcatgctgtaatccagcacttttgggggcgtgaggtggcggaatcatgaga 2700
Db 2641 ACGGTGGCTCATGCCCTGTAAATCCACGACTTTGGGGGGCTGAGTGGCGGAATCATGAGA 2700
QY 2701 tcaggagttcgagaccagcgtgcccagcagtggtgaaacccccctctctactaaaaatacaa 2760
Db 2701 TCAGGAGTTTCGAGACAGCAGCCTGGCCAGCATGTTGAAACCCCATCTCTACTAAAAATACAA 2760
QY 2761 aaaaatagctggcggtgtagcggggcgccgtgtaattcccagatactcaggaggtgaggta 2820
Db 2761 AAAATTAGCTGGCGGTGGTGACGGGGCCCTGTAAATCCCAGATACTCAGGAGGTGAGGTA 2820
QY 2821 ggaatacactgaaacctggaggtggaagtgtcagtgaaaccaagatcacgcgcaactgcac 2880
Db 2821 GGAGATCACTTGAACCTGGGAGGTGGAAGTGTCACTGTAACCAAGATCATCGCCACTGAC 2880
QY 2881 tccagcctgggagatggagcagactccaaactcaaaaaaataaaaaa 2930
Db 2881 TCCAGCCTGGCGATGGAGCGAGACTCCAACCTCAAAAAAATAAAAAA 2930

RESULT 2

AF244577 2930 bp mRNA linear PRI 06-AUG-2000
LOCUS Homo sapiens membrane glycoprotein HP59 (HP59) mRNA, complete cds.
DEFINITION AF244577
ACCESSION AF244577
VERSION AF244577.1 GI:9719373
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2930)
AUTHORS Fu,C., Bardhan,S., Cetateanu,N.D., Lloyd,S.R., Yan,H.-P.,
Carter,C.E., Shi,E., Venkov,C., Yakes,M.F., Page,D.L. and
Hellerqvist,C.G.
TITLE Identification of a novel membrane protein from mammalian cells
that interacts with the anti-pathoangiogenic compound CM101
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 2930)
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2000) Biochemistry, Vanderbilt University, School
of Medicine, 23rd Pierce, Nashville, TN 37232-0146, USA
FEATURES
Location/Qualifiers
1..2930
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..2930
/gene="HP59"
263..1873
/gene="HP59"
/note="similar to Na+ inorganic phosphate transporter"
/codon_start=1
/product="membrane glycoprotein HP59"
/protein_id="AAF97769.1"
/db_xref="GI:9719374"
/translation="MAGAMTPRPVQPARPGFGLSGRRSLLCQVASTPAHVGMRS
PVRLARNDGEESTDRTPLLGPAARAEAPVCCSARYNLAILAFFGFFIVIALRVNLS
VALVDVDSNTTLEDNRTSKAPESHSAPIKVHNQTKGYQWDAEQWIGLGSFFYGY
ITITQPGGVASKIGKMLLGFILGTAVLTFTIAADLVGVLIVLRALEGLGEGY
TPPAMHMSWAPLERSKLLISYAGAQLGTVLSLPLSGIICYVMWTVYVFFGT
GIFWFLWLVSDTPOKHFSHVEKEYILLRNOLSSOKSPVWPVILKSLPLWA
IVAFNSYNWTFYTLTLTPYMKELRENQENGFSLSLPYLGSWLCMILSQADN
LRAKNWSTLCVRRIFSLIGMIGPAVFLVAGFICGDSLAVAFITSTITLGGFCSSG
FSINHLDIAPSYAGILLTGITFTATPGMGPVIAKSLTPDNTVGEWQTVFYIAAII
VFGATFFTLFAKGEVQNNALNDHGRH"
BASE COUNT 729 a 698 c 681 g 822 t
ORIGIN

Query Match 100.0%; Score 2930; DB 9; Length 2930;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2930; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gttcgggtgaagccctccctcttaattatgtgcaattcaagtcctcccaactgccccgcgcgcaa 60
Db 1 GTTCGGTCCGAAGCCCTCCCTTTAATTTATGTGCAATTCAGATCCCTCCCTCCCGCCGCAA 60
QY 61 gcccccactcatctcgctgcggcaggggtgcccctcgacattacaagggggtgcagga 120
Db 61 GCCCCCACTCATCTCGCTGCGGGCAGGGTGGCCCTGTCACATTACAAAGGGGTGCAGGA 120
QY 121 gggggagagcgtgctccgaacacccggctcccccggcgtcgtagaccggcgggcgagcg 180
Db 121 GCGGGAGAGCGGTGCTCCGAACACCGGCTCCCGGCATGCGTAGACCGCGGGCGGAGCGG 180
QY 181 gctcaatttcgcgcacatcctacagagaactccccagaactccggttccctagtcacacccaa 240
Db 181 GCTCACTTTGCGCAATCTCTACGAGAACTCCCAAGAACTCCGCTTCCCTAGTCAACCCAA 240
QY 241 gccagagttgcccacacctaagatggcgggggcgatgacaccccccgcggtcca 300
Db 241 GCCAGAGTTGCCCAACACCTAAGATGCGGGGGGGCGATGACACCCCGCCCGCGGTCCA 300

QY 301 gccagctcgccgggcttgcggtctcgggcgccgctccctctctgcagggtgc 360
DB 301 GCCAGCTCGCGCGGGGCTCGGGCTGTCGGCGCGGCTCCCTCTCTCGCCAGGTGC 360
QY 361 gagtacacctgctcacgtagcgtcatgaggtctccgggttcgagacctgcccgggaacga 420
DB 361 GAGTACACCTGCTCACGTAGCGCTCATGAGGTCTCCGGTTCGAGACCTGCCCGGAACGA 420
QY 421 tggcgaagagcaagcagccagccctctctacggcgcccccagccgagccgc 480
DB 421 TGGCGAGAGAGCAGCAGCCAGCCCTCTCTACCGGGCGCCCGCAGGCGCGGAGCCGC 480
QY 481 tccagtgctgctcgtcgtctacaaacttagcaatttggcccttttttggttcttcac 540
DB 481 TCCAGTGTGCTCTGCTCTGTTACAACTTAGCAATTTTGGCTTTTGGTTCTTCAT 540
QY 541 tgtatgcatcagctgtaactgaggtgctgctgagtgtagtgatgtagatcaaatatc 600
DB 541 TGTGTATGCAATACGTGTGAATCTGAGTGTGCGTTAGTGGATATGGTAGATTCAAAATC 600
QY 601 aacttagaagataaagaacttccaagcgctgccagagcattctgctcccaataaagt 660
DB 601 AACTTTAGAGATTAATAGAACTTCCAAGCGGTGCCAGACATCTGCTCCCCATAAAGT 660
QY 661 tcatcataaataaacgggtgaagtagtaccatgggatgcagaaactcaagatggattct 720
DB 661 TCATCATAAATCAACGGGTGAAGTAGTACCAATGGGATGCAGAACTCAAGGATGGATTCT 720
QY 721 cgggttcctttttatggtacatcatcacagagattcctggaggatatgttgcagcaa 780
DB 721 CGGPTCCCTTTTATGGCTACATCATCACAGATTCCTGGAGGATATGTTGCCAGCAA 780
QY 781 aatagggggaaatgctctagatttggatcccttggcaactgctcctccaccctgtt 840
DB 781 AATAGGGGGAAATGCTGCTAGATTGGGATCCCTTGGCACGTCTCTCCACCCTGTT 840
QY 841 cactcccatctgctcagatttagagttggaccactcaattgtactcagagcactagaagg 900
DB 841 CACTCCCATTTGCTGCAGATTTAGAGTTGAGAGTTGGACCACATCATTTGACTCAGAGCAGTAGAAGG 900
QY 901 actagagaggggttacatttccagccatgcagtcgcatgtggttcttctgggtcccc 960
DB 901 ACTAGGAGAGGGGTTCACATTTCCAGCCATGCATGCCATGTGCTCTTCTTGGGCTCCCC 960
QY 961 tcttgaagaagcaaacctcttagcatttcgtatgcagagcagcagcttgggacagtaat 1020
DB 961 TCTTGAAGAAGCAAACTCTTAGCATTTGCTATGCAGAGGACACAGCTTGGGACAGTAAT 1020
QY 1021 ttctctctctcttcttggaaataattgctactatatagaattggacttatgtcttactt 1080
DB 1021 TTCTCTTCTCTTCTTGGAAATAATTTGCTACTATATGAATTTGGACTTATGCTCTTACTT 1080
QY 1081 ttttggactattggaataatttgggttcttcttggatcgtttagttagtgacacacc 1140
DB 1081 TTTTGGTACTATTGGAAATAATTTGGTTCCTTTTGGATCTGGTTCAGTTCAGTTCAGTTC 1140
QY 1141 acaaaaacacaagagaatttcccatatgaagaaggaaatacatcttctcatcaataagaaa 1200
DB 1141 ACAAAAACACAAGAGAAATTTCCCATATATGAAGAAGGAATACATTTCTTCATCATAGAAA 1200
QY 1201 tcagcttcttcacagaagtcagtcgctgggttacccattttaaaatccctgcacatttg 1260
DB 1201 TCAGCTTCTTTCACAGAAGTTCAGTGGCGTGGGTACCCATTTTAAATTCCTGCCACTTTG 1260
QY 1261 ggtatcgtagttgacactttcttcaactggaacttttttatactttattgacattatt 1320
DB 1261 GGTATCGTAGTTGACACTTTTCTTACAACTGGACTTTTATATCTTTATATGACATTAAT 1320
QY 1321 gctacttataagaagagacttaaggttcaatgttcaagagaatgggttttttattcttc 1380
DB 1321 GCCTACTTATAGAGGAGATCCTAGGTTCAATGTTCAAGAGAATGGGTTTTTATCTTC 1380
QY 1381 attgcctatttaggtctcttggttatgtatgatacctctgctggtcgaagctgtagacaattt 1440

DB 1381 ATTGCCTATTATTAGGCTTTGGTTATGATGATCCTGTGGTCAAGCTGCTGACAATTT 1440
QY 1441 aagggaataatgaaatttttcaacttatgtgtcagaagaatttttagccttatagaat 1500
DB 1441 AAGGGAAAATGAAATTTTTCAACTTTATGTGTTCAGAAATTTTATAGCCTTATAGGAAT 1500
QY 1501 gattgacactgagttattcctgtagctcgtgcttcaattggctgattattcttggc 1560
DB 1501 GATTGACCTTGCAGTATTTCTGTAGCTGTGGCTTCATTTGGCTGTGATTTATTTCTTTGGC 1560
QY 1561 cgtgtcttcttaactatatacaacaactgggagcgttttgcctcttcttgcgtttagcatt 1620
DB 1561 CGTGTCTTCTTAACATATATCAACAACACTGGGAGGCTTTTGTCTCTCTGGATTTAGCAT 1620
QY 1621 caacctctggatattgctcctctcgtatgctggtatcctcctcctgggcatcaacaatacatt 1680
DB 1621 CAACCATCTGGATATTGCTCTTCGTATGCTGTATCTCTCTGGCATCACAAATACATTT 1680
QY 1681 tgcactattccaggaatggttggccgctcattgctaaaagtctgacccctgataaacac 1740
DB 1681 TGGCATAATTCCAGGAATGGTTGGGCGCGTCATTGCTAAAAGTCTGACCCCTGATAACAC 1740
QY 1741 tgttgagaatggcaaacccgtgttctatatattgctgctgctattaaagtgttttgggtgccat 1800
DB 1741 TGTGTGAGAAATGCCAAACCGTGTCTTATATTGCTGTGCTATTAATGTTTTTGGTGCCAT 1800
QY 1801 tttctttacattatcccaagggtgaagtacaaaactgggctcctcaatgatcacccatgg 1860
DB 1801 TTTCTTTTACACTATTGCCAAAGGTGAAGTACAAAACACTGGGCTCTCAATGATCACCATGG 1860
QY 1861 acagacactgaaggaaccaaataataatcctgcctctattaaatgtatttttatttacc 1920
DB 1861 ACACAGACACTGAAGAACCAATAAATACTCTGCCTCTATTAATGTAATTTTATTTATC 1920
QY 1921 atgtaacctcaagtcctctctgtattgtgtgaagcattctatgtcttctttttaaattgtac 1980
DB 1921 ATGTAACCTCAAAAGTGCCTTCTGTATTGTGTAAGCATTTCTATGCTTTTTTTAAATTTGTAC 1980
QY 1981 tgttataatttttaagccctataatcaataataactgaaataactactagttccagaataaaaa 2040
DB 1981 TTGTATTAGATTTTAAAGGCTTATAATCATGAATAATACATAGTTGCCAGAAATAAATAA 2040
QY 2041 tgaactgtgttaattatgaataatgaatgaactagacttctacttttaggtttcacatc 2100
DB 2041 TGAACGTGTTTAAATTATGAATAATATGTAAGCTAGGACTTCTACTTTAGGTTTCATAC 2100
QY 2101 ctgctcgtcgtagtcgggaacatgaagtaggacagttctgttattttttagggccatact 2160
DB 2101 CTGCCCTGCTAGTCGGGCAACATGAAGTAGGACAGTTCTGTGATTTTTTATAGGGCCATACT 2160
QY 2161 aaaggaaatgagctgaacagacctcctcgtatccttttgccttaataaactagatgataat 2220
DB 2161 AAAGGAATGAGCTGAACAGACCTCCTGATACCTTTGCTTTAATTAACCTAGATGATAAT 2220
QY 2221 tctcaggtactgataaacacctgtgtgttcaacttctcctcataaaaaattgtcagctctc 2280
DB 2221 TCTCAGGTACTGATAAACACCTGTTGTTGTTCACTTTCTCTCATAAAAAATTTGTCAGCTCTC 2280
QY 2281 tctgacacttagacctcaaaccttttagcatctctgtgtgagctgacctccactgataattt 2340
DB 2281 TCTGACACTTAGACCTTCAAACTTTAGCATCTCTGTGGAGCTGCCATCCACTGCTGATAATTT 2340
QY 2341 cgcctggcaactggactgagggagtggtcccgagcagctgccaagcactcctccctcgtg 2400
DB 2341 CGCCTGGCAACTGGACTGAGGGAGGTGTCGCCAGGAGCTGCCAAGCACTCTCCCTCCCTGG 2400
QY 2401 cttcaggttcagagtgcccgagcgtttatcagagggagcagctccaaagcccagagcagtgctc 2460
DB 2401 CTTTCAGGGTCAGAGTGCCTCCAGCGTTTATCAGAGGCGAGCATCCAAGCCAGAGCCAGTGTC 2460
QY 2461 gactcttcggctgggtccttctcctctgaggggctatcaatgtgtgataaaagccctgact 2520


```
QY 2101 ctgcctctagtcggtgcaacatgaagtagcacagttctctgttgatttttagggccatact 2160
|||||
Db 1986 CTGCTCTAGTCGGCAACATGAGTAGCACAGTCTGTTGATTTTATAGGCCATFACT 2045
|||||
QY 2161 aaaggggaatgagctgaaacagacctcctgatacctttgcttaataaactagatgataat 2220
|||||
Db 2046 AAAGGGAATGAGCTGAACACAGCTCTGATACCTTTGCTTAATTAACATAGATGATAAT 2105
|||||
QY 2221 tctcaggtactgataaacacctgttgttgaacttctcctcataaaaattgtcagctctc 2280
|||||
Db 2106 TCTCAGGTACTGATAAACACCTGTTGTTGTTCACTTTCCTCATAAAAATGTGCAGCTCTC 2165
|||||
QY 2281 tctgacacttagacctcaaaacttagcatctctgtgagctgctccatccactgataattt 2340
|||||
Db 2166 TCTGACACTTAGACCTCAAACTTTAGCACTCTCTGTGGAGCTGCCATCCACTGTATTAATTT 2225
|||||
QY 2341 cgctggcaactggactggaggggagtggtgcccaggcagctgccaaactcctccctcctgg 2400
|||||
Db 2226 CGCCTGGCAACTGGACTGAGGGAGGTGTGCCAGGCAGCTGCCAAGCACTTCCCTCCCTGG 2285
|||||
QY 2401 cttcagggtcagagtcgccagctttatcagagcagcatccaagcccagagccagtgctc 2460
|||||
Db 2286 CTTCAGGGTCAGAGTGCCTGAGCGCTTTATCAGAGGCAGCATCCAAGGCCAGCCAGCTGTC 2345
|||||
QY 2461 gactcttcggctggtgctcttcctctgaggggctatcaatgtgtagataaagccctgagt 2520
|||||
Db 2346 GACTCTTCGGCTGTCCTCTTCCTCTGAGGGCTATCAATGTGTAGATAAAGCCCTGAGT 2405
|||||
QY 2521 agcagagcagtgagatccactgctatggtcttgatatacctcctcctccctccccc 2580
|||||
Db 2406 AGCACAAGACAGTGAATCCACTGCATGCTGTGTGATACATCTCAAACTTTCCCTTCCG 2465
|||||
QY 2581 agcacagagaaatctggctggcatcgaacctgcacaaagaaataatgc 2627
|||||
Db 2466 AGCACAGAGGAATATTGGCTGGCATGCACCTGCACCTGCACAAAGAAAATGC 2512
|||||

RESULT 5
HSA387747
LOCUS
DEFINITION Homo sapiens mRNA for sialin.
ACCESSION AJ387747
VERSION AJ387747.1 GI:6562532
KEYWORDS sialin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2512)
Verheijen,F.W., Verbeek,E., Aula,N., Beerens,C.E., Havelaar,A.C.,
Joosse,M., Peltonen,L., Aula,P., Galjaard,H., van der Spek,P.J. and
Mancini,G.M.
A new gene, encoding an anion transporter, is mutated in sialic
acid storage diseases
Nat. Genet. 23 (4), 462-465 (1999)
20047778
2 (bases 1 to 2512)
Verheijen,F.W.
Direct Submission
Submitted (03-JUN-1999) F.W. Verheijen, Erasmus University, Dept.
Clinical Genetics, P.O. Box 1738, 3000 DR Rotterdam, NETHERLANDS
Location/Qualifiers
1..2512
/organism="Homo sapiens"
/db_xref="taxon:9606"
271..1758
/codon_start=1
/product="sialin"
/protein_id="CAB62540.1"
/db_xref="GI:6562533"
/translation="MRSPPVRLARNDGESTDRTPLLPGAPRAEAPVCCSARNLAI
LAFFGFFIVIALRVNLSVALVDMVDSNTTLEDNRTSKACPESAPIKVHHNOTGKKYQ
```

```
WDAETQGWILGSFFYYIITQIPGGYVASKIGKMLGFIGLITAVLFLFTPIAADLG
VGPLIVLRALGEGVTFPAMHAMWSWAPLERSKLISISYAGOLGTWISLPISG
IICYYMNTVYVFEPTGIFWELLWLWSDTPQKHKRI SHYEXEYTLSSIRNQLSS
OKSPVWPILAKSLPLAWAIVAHFSYNTFFYLLPILPYMKELIRFNQVNGFLSSLP
YLGSMILSGQAADNLARKWNFSTLCVRRIFSLIGMIGFAVELVAAGF1GCDTSLA
VAFITISTLLGFCSSGSFINHLDIAPSYAGILGILITNTFATIPGMWGPVIAKSLTPD
NTVGEWQTVFYIAAAINVFAGIFFTLFAKEGVQNWALNDHHGHRH"

BASE COUNT 612 a 585 c 566 g 749 t
ORIGIN

Query Match 80.5%; Score 2359.4; DB 9; Length 2512;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 2437; Conservative 0; Mismatches 6; Indels 64; Gaps 1;

QY 185 actttgcgcaactctctacagaaactcccaagaactccgcttccctagtccaacccaagcaa 244
|||||
Db 6 ACTTTGCGCAATCTCTACGAGAACTCCCAAGAACTCCGCTTCCTCTAGTCCAACCCAAGCA 65
|||||
QY 245 gqgttgccacacctaagatggcgggggg-----cgatgacacccgccccccgggtcca 275
|||||
Db 66 GAGTTGCCACACCTAAGATGGCGGGGGGGGGGGGCGCGGCGGCGGCGGCGGCGGCGG 125
|||||
QY 276 -----cgatgacacccgccccccgggtcca 300
|||||
Db 126 GACCGCGGGGACTAGACGTGGCGGGCGGCGGTGTCTATCGCCCGCGCGCGCGGTCCA 185
|||||
QY 301 gccagctggccccgggggcttcgggctgtcggggcgggcgctccctctctgcaggtggc 360
|||||
Db 186 GGCAGCTCGCGCGGGGGGCTTCGGGCTGTGCGGCGGGGCGGCTCCCTTCTTGCACAGTGGC 245
|||||
QY 361 gqgtacacctgtcctgtagcgctcaggtctcctcgggttcgagacctggcgccgaaagca 420
|||||
Db 246 GAGTACACCTGTCTACGTAGGCGTCTAGAGGTCTCGGTTTCGAGACCTGGCGCGGAACA 305
|||||
QY 421 tggcgagagagcagcggaccgcacgcctctctacccggcgccccacagggcgccgaagccgc 480
|||||
Db 306 TGGCGAGGAGAGCAGCGGACCGCACGCCCTCTTCTACGCGGCGCGCGCGCGGCGGAAGCGCG 365
|||||
QY 481 tccagtgctgctctgctcgttacaacttgcatacttgccttttttggcttttcttctcat 540
|||||
Db 366 TCCAGTGTGCTGCTCTGCTGCTGTACAACTTAGCAATTTTGGGCTTTTGGTGTCTTCTCAT 425
|||||
QY 541 tgttatgcattacgtgtgaatctgagttgctgttagtggtatggttagattcaaatatc 600
|||||
Db 426 TGTGTATGCAATTACGTGTGAATCTGAGTGTGCGTTAGTGATATGGTAGATTCAATATC 485
|||||
QY 601 aactttaagaataatagaacttccaaaggcgtgtccagagcattctgtccataaaagt 660
|||||
Db 486 AACTTTAGAAAGATAATAGAACTTCCAAGGCGGTGTCAGAGCACTTCTGCTCCCATAAAAGT 545
|||||
QY 661 tcatcataatcaaacgggtaagaactacaatggatgcagaaactcaagatgatctct 720
|||||
Db 546 TCATCATATCAAAACGGGTAAAGATACCAATGGGATGCAGAAACTCAAGGATGGATTCT 605
|||||
QY 721 cggttccttttttatggctacatcatcacacagattcctgagagatatgttgccagcaa 780
|||||
Db 606 CGGTTCTTTTATGGCTACATCATCACACAGATTCTCTGGAGGATATGTTGCCAGCAA 665
|||||
QY 781 aaatgggggggaaaaatcgtgctaggatttgggaccttggcactgtgctcctaccctgtt 840
|||||
Db 666 AATAGGGGGGAAAAATGCTGCTAGGATTTGGGATCCTTGGCACTGTCTCTCACCCCTGTT 725
|||||
QY 841 cactcccatgtcagatttaggagttggaccactcattgtactcagacactagaagg 900
|||||
Db 726 CACTCCCATGTCTGAGATTTAGGAGTTGGAGACTCATTCATTCTACAGAGCACTAGAAG 785
|||||
QY 901 actaggagagggtgtttacatttccagccatgcattgccattgtgtctcttgggctcccc 960
|||||
Db 786 ACTAGGAGAGGGTGTGTACATTTCCAGCCATGCATGCCATGTGCTTCTTCTTGGGCTCCCC 845
|||||
QY 961 tctttaaagaagcaaaccttcttagcatttctgatatcagagagcacagcttggagacagtaat 1020
|||||
```

| | | | | | | | | |
|----|------|-----|------------------------|---------------------------------|-----------------------------------|----------------|------------|-----|
| Db | 846 | TC | TTGAAAGAACAACTTC | TCTAGACATTC | TATATG | CCAGGACAGCAGCT | TGGCAGTAAT | 905 |
| Qy | 1021 | tt | ctctctctctcttggaataatt | tgctactatatgaatt | ggacttatgtctctctactt | | 1080 | |
| Db | 906 | TT | CTCTTCCTCTTCTTGGAATAAT | TGCTACTATATGAAT | TGGACTTATGCTTCTCTACTT | 965 | | |
| Qy | 1081 | tt | tgtgactattggaataatt | tgtgttctctt | tgtggaatctggttagttagtgaacacc | 1140 | | |
| Db | 966 | TT | TGTGTACTATTGGAATAAT | TTTGTGTCTTTTGTGTGATCT | TGGTATGTAGTAGTGCACACC | 1025 | | |
| Qy | 1141 | ac | aaacacacagagaattt | cccatatgaagaagaac | atcatcttccatcataaagaaa | 1200 | | |
| Db | 1026 | AC | AAACACACAGAGAAT | TTCCATATGAAGAAGATACAT | TTCTTCATCATTAAGAAA | 1085 | | |
| Qy | 1201 | tc | agcttcttcacagaagtc | agtcgctgggtaccacatt | ttaaaaaccctgccactttg | 1260 | | |
| Db | 1086 | TC | AGCTTTCTTCACAGAAGTC | AGTCGCTGGGTACCCAT | TTTAAATCCCTGCCACTTTG | 1145 | | |
| Qy | 1261 | gg | ctatcgtagttgcacactt | tcttacaactggactt | tttatacttttattgacattatt | 1320 | | |
| Db | 1146 | GG | TATCGTAGTTGTCACACT | TTTCTTACAACTTGGACT | TTTTTATACTTTATTGACATTAAT | 1205 | | |
| Qy | 1321 | gc | ctactatataagagagatc | ctaagctcaatgtccaagata | gggttttttatcttc | 1380 | | |
| Db | 1206 | GC | TACTATTATGAAGGAGATC | TAAAGTTCAAATGTT | CAAGAGAAATGGGTTTTTATCTTC | 1265 | | |
| Qy | 1381 | att | gcctatttagcctcttggt | tatgatcctgctgtgcaagctgcgcacaatt | 1440 | | | |
| Db | 1266 | AT | TGCCCTATTAGGCTCT | TGGTTATGATGATCTGTCT | GGTCAGCTGCTGACAAATTT | 1325 | | |
| Qy | 1441 | a | gggcaaaatggaattt | tcaactttatgtgtgcgagaatt | ttttagccttataggaaat | 1500 | | |
| Db | 1326 | A | AGGGAAAAATGGAAT | TTTTTCAACTTTATGTT | CGCAGAAATTTTAGCCCTTATAGGAAT | 1385 | | |
| Qy | 1501 | ga | ttagcactgcagtatct | octagtagctgcgtctcaatt | ggctgtgatattctcttggc | 1560 | | |
| Db | 1386 | GA | TGGACCTGTCAGTAT | TCTCGTAGCTGCTGGCTTCAT | TGGCTGTGATTAATCTTTTGGC | 1445 | | |
| Qy | 1561 | cg | tgtctctcaactatata | caacaacactggagcgctt | gtctctctctggaattagcat | 1620 | | |
| Db | 1446 | CG | TGCTTTCCTPAACCTAT | ATCAACACACTGGGAGGCT | TTTTTGCCTCTCTGGAATTTACAT | 1505 | | |
| Qy | 1621 | ca | accatctggatattgctc | ctctcgtagtgcgtatcct | cctgggcacatcaaaatacatt | 1680 | | |
| Db | 1506 | CA | ACCATCTGGATATTGCT | CTCTGCTATGCTGGTAT | CTCTCGTGGCATCACAAATACATT | 1565 | | |
| Qy | 1681 | tg | ccactattccaaggaat | gggtgggcccgtcat | tgtcaaaagctgaccctcatacac | 1740 | | |
| Db | 1566 | TG | CCACTATTCCAGGAAT | GGTTGGGCCGCTCAT | TGCTAAAAAGTCTGACCCCTGATAACAC | 1625 | | |
| Qy | 1741 | tgt | gagaatggcaaacgct | tctatattgctgctgctatt | aatgtttttagtgcatt | 1800 | | |
| Db | 1626 | TG | TGGAGAAATGCCAAACCG | TGTCTATATGCTGCTGCTAT | TATGTTTTTGTGGTGGCAT | 1685 | | |
| Qy | 1801 | tt | ctttcacactattcc | caaaaggtagaagta | caaaacctggcctctcaatgatcaccattg | 1860 | | |
| Db | 1686 | TT | CTTTTACACTATTGCC | CAAAAGTGAGTACAAAC | TGGGCTCTCAATGATCACCATTG | 1745 | | |
| Qy | 1861 | ac | acagacactgaagga | acaaataaactc | gtcctctataatgtatttttattatc | 1920 | | |
| Db | 1746 | AC | ACACACACTGAAGGAAC | CAATAAATCTCGCTCTCT | TATTAATGTATTTTTTATTTATC | 1805 | | |
| Qy | 1921 | at | gaacctcaagagctct | tgattgtgaagcattct | ctcttttttaattgtac | 1980 | | |
| Db | 1806 | AT | GTAAACCTCAAAAGT | GCCCTCTGTATGTTGTAAGCAT | TCTATGCTTTTTTTTAAATGTATC | 1865 | | |
| Qy | 1981 | tt | gtattagatttt | taagccctataatcat | gaaatcacactagttgccagaataataaaa | 2040 | | |
| Db | 1866 | TT | GTATTAGATTTTAA | AGGCTTAAATCAT | GAAATATCACTAGTTGCCAGAAATAATAAA | 1925 | | |
| Qy | 2041 | t | gaactgtgttaatt | atgaataatgt | taagcttaggactctctacttttaggtcacatc | 2100 | | |
| Db | 1926 | TG | AACTGTGTTTAAAT | TATGAATAATATGAT | GAAGCTAGGACTTCTACTTTAGGTTTACATAC | 1985 | | |

| | | | |
|------------|---|---|---------|
| QY | 2101 | ctgcctgtagtgcgggcaacatgaagtaggacagtctctgtgatttttttagggccaatact | 2160 |
| Db | 1986 | CTGCCTGCTAGTCGGGCAACATGAAGTAGGACAGTCTGTGTGATTTTTTTAGGGGCATACT | 2045 |
| QY | 2161 | aaagggaatgagctgaacagacctcctgatcacctcttgcttaataaactagatgataat | 2220 |
| Db | 2046 | AAAGGGAATGAGCTGAACAGACCTCCTGTATACCTTTGCTTAAATAACTAGATGATAAT | 2105 |
| QY | 2221 | tctcaggtactgataaacacacctgtgtgttcttcacttctctcataaaaaattgcagctctc | 2280 |
| Db | 2106 | TCTCAGGTACTGATAAACACACCTGTGTGTGTTCACTTTCTCTCATAAAAATTTGCAGCTCTC | 2165 |
| QY | 2281 | tctgaaccttagaacctcaaaactttagatcctctgtggagctgcacatcgtataattt | 2340 |
| Db | 2166 | TCTGACACTTAGACCTCAAACTTTAGCATCTCTGTGAGCTGCCATCCACCTGTATAATTT | 2225 |
| QY | 2341 | cgcttgcacaactgactgagggagtggtgccagcgagctgccaaagcactccctccctgg | 2400 |
| Db | 2226 | CGCTTGGCAACTGACTGAGGGGAGTGTGCCAGGCAGCTGCCAAGGACTCCCTTCCCTGG | 2285 |
| QY | 2401 | cttcagggtcagagtgcacacgctttatcatagaggcagcatccaaagccagagccagtgctc | 2460 |
| Db | 2286 | CTTCAGGCTCAGATGCCACGCTTTATCAGAGCAGCATCCAAGCCAGAGCCAGTGCTC | 2345 |
| QY | 2461 | gactcttcggtgctgtccttctcctcaggggctatcaatgtgtagataaagccctgaagt | 2520 |
| Db | 2346 | GACTCTTCGGCTGTGGCTCTTTCCTCTCAGGGGCTATCAATGTGTAGATAAAGCCCTGAGT | 2405 |
| QY | 2521 | aggcaagagcagtgagatccactgctggtcttgatcacatccctcaaaacttccctccg | 2580 |
| Db | 2406 | AGGCAAGAGCAGTGAGATCCACTGCTATGCTGTGTATACATCTCTCAAACTTTCCCTCCC | 2465 |
| QY | 2581 | agcacaggaataattggctgtggcctgcaacctgcaaacctgcaaaagaaaaatgc | 2627 |
| Db | 2466 | AGCACAGGAAATATTGGCTGGCATGCAACCTGCAAAAGAAAAATATGC | 2512 |
| RESULT | 6 | | |
| AF244578 | AF244578 | Ovis aries | 2844 bp |
| LOCUS | Ovis aries | membrane glycoprotein Sp55 (sp55) | linear |
| DEFINITION | Ovis aries | membrane glycoprotein Sp55 (sp55) | mrna |
| ACCESSION | AF244578 | | |
| VERSION | AF244578.1 | GI:9719375 | |
| KEYWORDS | sheep. | | |
| SOURCE | Ovis aries | | |
| ORGANISM | Ovis aries | | |
| REFERENCE | Fu,C., Bardhan,S., Cetateanu,N.D., Lloyd,S.R., Yan,H.-P., | | |
| AUTHORS | Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; | | |
| | Bovidae; Caprinae; Ovis. | | |
| TITLE | 1. (bases 1 to 2844) | | |
| JOURNAL | Fu,C., Bardhan,S., Cetateanu,N.D., Lloyd,S.R., Yan,H.-P., | | |
| REFERENCE | 2. (bases 1 to 2844) | | |
| AUTHORS | Fu,C., Bardhan,S., Cetateanu,N.D., Lloyd,S.R. and HELLERQVIST,C.G. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (13-MAR-2000) Biochemistry, Vanderbilt University, School | | |
| | of Medicine, 23rd Pierce, Nashville, TN 37232-0146, USA | | |
| FEATURES | Location/Qualifiers | | |
| source | 1. .2844 | | |
| | /organism="Ovis aries" | | |
| | /db_xref="taxon:9940" | | |
| | /cell_type="endothelial cell" | | |
| | /tissue_type="lung" | | |
| gene | 1. .2844 | | |
| | /gene="sp55" | | |
| CDS | 84..1571 | | |
| | /gene="sp55" | | |
| | /note="similar to Na+ inorganic phosphate transporter" | | |
| | /codon_start=1 | | |

| | | | |
|------------|--|--|----------------------------|
| Db | 1864 | ACCATACCTAAGAGATGAGCTGAATAGACACCTTCTATACCTTTGCTTAATTAAGGTGG | 1923 |
| Qy | 2213 | atgataattctcaggtactagataaacacacgtgtgtgtgttcaacttctcctataaaa----- | 2267 |
| Db | 1924 | ATAATAATCTCAGGTCTTGTAAACATCTGTTTGTACACCTTCTCTCAAAAATATT | 1983 |
| Qy | 2268 | -attgtcagctctctctgacacacttagacctcgaacatttagcatctctgtgagctccat | 2326 |
| Db | 1984 | TGTCATCAGCAATPCCCTGACATGAGTGTCTCAAACTTTAGCCTCTCCACGAGCTGGCAG | 2043 |
| Qy | 2327 | ccactgataatttcctcgcgaactggactgagggagtggtgccaggcagctgccaag | 2386 |
| Db | 2044 | CCACTGTATCATTTACCTTGGCAACTTTCACGTAGGGAAGCATCCCGAGCAGCTGCACA | 2103 |
| Qy | 2387 | cactcctcctcgtgttcagggttcagagtgccagcggtttatcatcagaggcagcatccaagc | 2446 |
| Db | 2104 | TGTCCTCTCTGTGCTTCAGGACAGTGGCCACACTT-----AGGCAGCATCCAAGA | 2156 |
| Qy | 2447 | ccaggcaggtctcgtactctcgtcgtgtggtccttctcctctgaggggtatcaatgtgtag | 2506 |
| Db | 2157 | CCAGGCTCAGCGCAAGGCTTTGACGGTATTCTTCCCTT--GGGGCTGTTAATGTGTGG | 2214 |
| Qy | 2507 | ataaagcctcagtaggcaagcagtgagatccatccatgctctatgtcttgatacatctca | 2566 |
| Db | 2215 | ATGAAGCCTTGAGCCCAACAGGACAGCGCATCCAGATCATGGTTTCCATGACCCCTC- | 2273 |
| Qy | 2567 | aactttccttcccagcacagaggaattattgctggcatgcaacctgcgaaaagaaa | 2622 |
| Db | 2274 | TCCCTTCCCTTCCAGCACACTGAGTATTGCTGSCATGTAACTTGCAAAAGAAA | 2329 |
| RESULT | 7 | | |
| AX207626 | | | |
| LOCUS | AX207626 | 2844 bp | DNA linear PAT 31-AUG-2001 |
| DEFINITION | Sequence 3 from Patent WO0156598. | | |
| ACCESSION | AX207626 | | |
| VERSION | AX207626.1 | GI:15422331 | |
| KEYWORDS | Ovis sp. | | |
| SOURCE | Ovis sp. | | |
| ORGANISM | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Caprinae; Ovis. | | |
| REFERENCE | 1 (bases 1 to 2844) | | |
| AUTHORS | Hellergvist, C.G. | | |
| TITLE | Methods for preventing or attenuating pathoangiogenic conditions by using the gbs-toxin (cm101) receptor as a vaccine | | |
| JOURNAL | Patent: WO 0156598-A 3 09-AUG-2001; | | |
| FEATURES | VANDERBILT UNIVERSITY (US) | | |
| source | Location/Qualifiers | | |
| | 1..2844 | | |
| | /organism="Ovis sp." | | |
| | /db_xref="taxon:9939" | | |
| CDS | 84..1571 | | |
| | /note="unnamed protein product" | | |
| | /codon_start=1 | | |
| | /protein_id="CAC60359.1" | | |
| | /db_xref="GI:15422332" | | |
| | /translation="MKSPVSDLPDSGEGSDRTPILORAPRAEPAPVCCSARYNLAF | | |
| | LSFGFFVLYLSRVNLVALVDMDSNTAKDNRTSYEAEHSAPIKVLHNOTGKKYR | | |
| | WDAETQGLGFFYGIITQIPGGTVASRSGKLLGLGFIATFIATFTLPDADF | | |
| | VGALVALRLEGVGVTPAMHMSWAPLERSKLLSIYAGAQLGVVSLPLSG | | |
| | VICYNNNTVYFFYGIWIWFIWICIVSDTPETHKTITPYREYILSKNLQSS | | |
| | OKSPVIMPLKSLPWLAIWVAHFSYNTWFTLLTLLPTVMKVLRFNQENGLSAVP | | |
| | YLGCWLMTLSQAADNLRARNFSTLWVRVFSLIGMTGPAIFLVAAGFICDYSIA | | |
| | VAFUITSITLGGFCSSGSINHLDIAPSAGILLGITNTFTATIPGMGPIIARSITPE | | |
| | NTIGBWQVFCIAAINVFGAIFFTLFKAGKGVQNNALSDHQHRN" | | |
| BASE COUNT | 745 a | 623 c | 594 g 882 t |
| ORIGIN | | | |

Query Match 53.6%; Score 1570.8; DB 6; Length 2844;
Best Local Similarity 82.0%; Pred. No. 0;

| | | | |
|--|------|---|------|
| Matches 1916; Conservative 0; Mismatches 387; Indels 33; Gaps 8; | | | |
| QY | 309 | ggccgggggcttcgggctgcggcgccgctcctctctccaggtgagctacac | 368 |
| Db | 5 | GGGCGGGGGGCTTCGGCGGTCCGCTGGAGCTCTCTTTTCCGGGAGCAGGTTTGGCCG | 64 |
| QY | 369 | ct--gtcacgtaggcgtcatgatggtctccggttcgagacctggcccggaacgatggcga | 426 |
| Db | 65 | GTAGCTCCCTGAAGGCATCATGAAGTCCCGGTTTCGGACTTAGCCCCGAGCAGCGCA | 124 |
| QY | 427 | ggagagcagcagcgcgcctctctacoggcgcccaacggcggaagccgctccagt | 486 |
| Db | 125 | GGAGGGCTCGGACCGCACACCCGCTCTGACGCGCCCCCGCGGCGGAACCCGCTCCAGT | 184 |
| QY | 487 | gtcgtcctcgtcgttacaaacttagcaatttggccttttttgggtttctctcattgtga | 546 |
| Db | 185 | ATGCTGCTCTGCTGTGTACAACTAGCATTTTGTCTTTTGGTTTCTTCTGTTCTCTA | 244 |
| QY | 547 | tgcattacgtgtgaatctgagtggtcgttagtggtatggttagattcaatacaacttt | 606 |
| Db | 245 | TTCAATTACGGGTGAATCTGAGCGTTGCACCTAGTGGACATGGTGGATTCAAACACA | 304 |
| QY | 607 | agaagataatagaacttccaagcggtgccagagcattctgccataaaagtcatca | 666 |
| Db | 305 | CAAAAGATTAATAGAACGCTCTTACGAGTGTGCAGAGCATTTGCTCCCATAAAGTCTTCA | 364 |
| QY | 667 | taatcaaacgggttaagaagtaccatgggatgcagaaactcaagatggattctcggttc | 726 |
| Db | 365 | CACCAAAACGGGTAAAAAGTACCGTGGGATGCAGAACTCAAGGATGGATTCTCGGATC | 424 |
| QY | 727 | cttttttattggtcatcatcacagattcctggaggagatatgttgcagcaaaaatagg | 786 |
| Db | 425 | TTTTTTCTATGCTACATCATCACAAATTCCTGGAGGATATGTTGCCAGCAGAAAGTG | 484 |
| QY | 787 | ggggaataatgctcaggtattgggatccctcctgcaactgctcaccctgttccactcc | 846 |
| Db | 485 | GGGGAAGCTGTGCTAGGATTCGGGATCTTTGTGACAGCTATCTTCACTTCACTCTCC | 544 |
| QY | 847 | cattgctcagatttaggagttgaccactcattgtactcagagcactagaagactagg | 906 |
| Db | 545 | CCTCGCTGCAGATTTTCGGAGTCGGAGCCCTTGTTCACCTCAGGGCACTAGAAGGGCTAGG | 604 |
| QY | 907 | agaggtgtttacatttccagccatgcacgtggtccttcttgggtccccctcttga | 966 |
| Db | 605 | AGAGGTTGTACATATCCAGCCATGCATGCTGTGCTTTCATGGGCTCCCTCTCTGA | 664 |
| QY | 967 | agaagcaaaactcttagcaatttcgtatgcagagcacacgttggacagatattctct | 1026 |
| Db | 665 | AAGAAGCAAGCTTCTGAGTATTTTCATATGAGGAGCACAACCTGGGACAGTAGTTCCTCT | 724 |
| QY | 1027 | tctcttcttggaataattgctactatatgaattggacttatgcttactctttttgg | 1086 |
| Db | 725 | TCCTCTTCTGGAGTAATTTGCTACTATATGAATTGGACTTATGCTCTTCTATTCTTTGG | 784 |
| QY | 1087 | tactatggaataatttgggttcttcttggatcctggttagtgtagtagtagtagtag | 1146 |
| Db | 785 | CATTGTTGGAATCATCTGCTTTATTTTATGGATCTGCTAGTAGTATACACGAGAAAC | 844 |
| QY | 1147 | acacagagaatttccattatgaaaagaatacatcttctcattcattagaatacagct | 1206 |
| Db | 845 | TCACAAGACATCACCTCCCTTATGAAAGGAGTATATCTTTTCATCATTAATAAATCACT | 904 |
| QY | 1207 | ttctcacagaactcagtcgctgggtaccocattttaaataccctgccactttgggctat | 1266 |
| Db | 905 | CTCTTCACAGAACTAGTCCGCTGGATACCTATGCTGAATCACTGCCACTTTGGGCTAT | 964 |
| QY | 1267 | cgtagttgcacactttcttacaactggactttttaactttttagacattattgacctac | 1326 |
| Db | 965 | TGTCGTTGCACATTTTCTTACAACCTGGACTTTTATATCTTGTGTGACCTTATTGCTAC | 1024 |
| QY | 1327 | ttatatgaaggagatcctcaaggttcaatgttcaagagaatgggttttttatcttcaatgccc | 1386 |
| Db | 1025 | TTACATGAGGAAGTCTCTAAGGTTCAATATTCAAGAGAATGGGTTTTTATCTGCGATCCC | 1084 |

| | | | |
|----|------|---|------|
| Qy | 1387 | ttatttaggccttggttatgtatgatacctgtctctggtctcaagctcgtcgacaatttaagggc | 1446 |
| Db | 1085 | TTATTTAGGTGTGTGTATGTATGATCATCTGTCGGGTCAAGCTGCTGCAAAATTTAAGGGC | 1144 |
| Qy | 1447 | aaatggaattttcaacittatgtgttcgcagaatttttagccttataggaatgattgg | 1506 |
| Db | 1145 | AAGATGGAATTTTTCAACTCTGTGGGTTCCGAAGAGTTTTTAGCCTTATAGGGATGATTGG | 1204 |
| Qy | 1507 | acctgcagtaattcctggttagctgctggcttcattgctgtgattattcttggcgcgttgc | 1566 |
| Db | 1205 | ACCTGCGATATTCTCTGGTTGCCGAGGATTTATAGSCTGTGATTATTCTCTTGGCTGTTCG | 1264 |
| Qy | 1567 | tttctaactatatacaaacactgggagcttttgcctctcttgatttagcatcaaccca | 1626 |
| Db | 1265 | ATTCTTAACCATATCAACAACCTCGGGAGGCTTTTGCTCTCTGGATTAGCATCAACCA | 1324 |
| Qy | 1627 | tcgggattcgtccctcgtatcgtcgttatcctcctggccatcacaaatcatttgcac | 1686 |
| Db | 1325 | TCYGGACATTCGTCCTCGTATGCTGGTATTTCTCTGGCATCAAAATACCTTTGGCAC | 1384 |
| Qy | 1687 | tattccagggaatggttgggcccgcctattgctaaaaagctcgacccctgataacactgttg | 1746 |
| Db | 1385 | TATTCTTGGAATGATTGGGCCCATCATTTGCCAGAAGCTTACCCTCGAGAACACTATTGG | 1444 |
| Qy | 1747 | agaatggcaaacogtatttotaattgctgctgctattaatgttttttggcgcaatttcctt | 1806 |
| Db | 1445 | AGAAATGGCAAACTGTTTCTGTCATCGCTGCTCTATCAATGTATTTGGTGGCATTTTCTT | 1504 |
| Qy | 1807 | tacactattcgccaaagtggaagtacaaaactggcctcctcaatgatcaccaatggacacag | 1866 |
| Db | 1505 | CACACTATTGCGCCAAAGGTGAAGTGCAAAACTGGGCCATCAGTGATCACCAAGGACACAG | 1564 |
| Qy | 1867 | acactgaaggaaaccaataaatactcgtcctctattaatgtatttttatttatcatgtaa | 1926 |
| Db | 1565 | AAACTGAAGGAACCAATAAATAATCTGTCTCTATTAAATGTATCTTTTATCATGTAA | 1624 |
| Qy | 1927 | cccaaaagtcctt----ctgtattgtgaagcattctatgt-----cttttttt | 1972 |
| Db | 1625 | CCTAAAGTGCCCTTGATATTTTAAATGTGAAGCAATCTATATACAAGATAAAATGTAC | 1684 |
| Qy | 1973 | aattgactctgtattagatttttaaggccctaatactgaagaatatcactagttgcccagaa | 2032 |
| Db | 1685 | TAGAAAAATTGTGTAGATTGTGAGGCTTGTAAATCATGAAATGTCACTAGTTCGCATAT | 1744 |
| Qy | 2033 | taataaaatgaactgtgtttaaattatgaataatatgtgaagctgagacttctactttaggt | 2092 |
| Db | 1745 | AAGCAAAATTAGCTATTTTTTAAATTTAATTAACCCCGTTGCTGGAACCTACAAATTCAGGG | 1804 |
| Qy | 2093 | tcacatacctgctcgtactcgggcaacatgaagtagagacagttctgttatttttagg | 2152 |
| Db | 1805 | TCACATATCTGGCTGCAAGTCAAGGCAACCCACAATAGGGGAGTGTCTATTTA-TTTATAAG | 1863 |
| Qy | 2153 | gcacataaagggaatgagctgaaacagacctcctgataccttctgttaataataaactag | 2212 |
| Db | 1864 | ACCATACCTAAAGAGATGAGCTGAATAGACCCCTTCATACCTTTGCTTAATTAAGGTGG | 1923 |
| Qy | 2213 | atgataattcgaagtactgataaacacotgttggttccacttctcctcatabaaa----- | 2267 |
| Db | 1924 | ATAATAATTCTCAGGCTTGTGTTTAAACATCTGTTTTTGTGTACACCTTCCTCAAAAATTAAT | 1983 |
| Qy | 2268 | -atgttcagctctctcgacacttagacctcaaaccttttagcatctcgtggagctgcacat | 2326 |
| Db | 1984 | TGTATCAGCAATCCCTGCACATGTAGGTCTCAAACTTTAGGCTCTCCACGGAGCTGGCAG | 2043 |
| Qy | 2327 | ccactgtataatttcctggccaactggactgagggagtgccccaggcagctgcgaag | 2386 |
| Db | 2044 | CCACTGTATCTACGCTTGCAACTTCACGTGAGGGAAGCATGCCAGGAGCTGCCACA | 2103 |
| Qy | 2387 | caactcctccctcgttctcaggctcagagtgtccccagctttatcacagaggcagcatccaagc | 2446 |
| Db | 2104 | TGTCCCTCTCTGTGGTTTCAGGACAGTCCCGCAGCACTT-----AGCAGCATPCCAAGA | 2156 |

[illegible]

* consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 35016: contig of 35016 bp in length
 * 35017 35116: gap of 100 bp
 * 35117 38963: contig of 3847 bp in length
 * 38964 39063: gap of 100 bp
 * 39064 56641: contig of 17578 bp in length.

FEATURES

source

Location/Qualifiers

1..56641
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="RP3-515P18"
 /clone_lib="RPCI-3"
 1..35016
 /note="assembly_fragment:00367
 clone_end:77
 vector_side:left"
 35117..38963
 /note="assembly_fragment:00264
 fragment_chain:1"
 39064..56641
 /note="assembly_fragment:00607
 fragment_chain:1"

misc_feature

misc_feature

misc_feature

BASE COUNT 16181 a 12335 c 12027 g 15895 t 203 others
 ORIGIN

Query Match 40.7%; Score 1193.6; DB 2; Length 56641;
 Best Local Similarity 99.3%; Pred. No. 3.le-271;
 Matches 1199; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1722 gctgacccctgataacacgttgagaaatggcaaacctgttctatatattgctgtgcta 1781
 |||||
 Db 43034 GTCCTCTTTATAGAACACTGTTGGAGATGGCAACCGTGTCTATATATGCTGCTA 43093
 |||||

QY 1782 ttaatgtttttggtgccattttcttacactattcgccaaagtggaagtacaaactggg 1841
 |||||

Db 43094 TTAATGTTTTTGGTGCCATTTCTTTACACTATTGCCAAAGTGAAATACAAAACGGG 43153
 |||||

QY 1842 ctctcaatgatcaccatgagacacacactgaagaaaccaataaataatcctgccttat 1901
 |||||

Db 43154 CTCTCAATGATCACCATGGACACAGACACTGAAGGAACCAATAAATAATCCTGCCCTAT 43213
 |||||

QY 1902 taatgtattttatttatcatgtaacctcaaatgagcttctgtattgtgaagcattcta 1961
 |||||

Db 43214 TAATGTATTTTATTATCATGTAACCTCAAGTGCCCTCTGTATTGTGTAGCATTCTA 43273
 |||||

QY 1962 tgccttttttaattgtactgttattagatttttaaggcctataatcatgaataatcact 2021
 |||||

Db 43274 TGCTTTTTTTTAAATGCTACTTGTATTAGATTTTAAAGGCTATAATCATGAATAATCACT 43333
 |||||

QY 2022 agttgccagaataataaaatgaactgtgttaattatgaataatgatgaagcttagact 2081
 |||||

Db 43334 AGTTGCCAGAAATAAATAATGAACCTGTGTTTAAATATGAATAATATGTAAGTAGACTT 43393
 |||||

QY 2082 ctacttttagttcacataccttgcctgctagtgcgggcaactgaagttagacagttctgt 2141
 |||||

Db 43394 CTACTTTAGTTTCATACACTCTGCTGCTAGTCTGGGCAACATGAAGTAGGACAGTTCTGT 43453
 |||||

QY 2142 gattttttagggccataactaagggaatgagctgaacagacctcctgatacctttgctt 2201
 |||||

Db 43454 GATTTTTTAGGGCCATACTAAAGGGAATGAGCTGAACAGACCTCTGTATACCTTTGCTT 43513
 |||||

QY 2202 aattaaactagatgataatctcaggtactgataaacacctgttgttgcactttccctc 2261
 |||||

Db 43514 AATTAACACTAGATGATAATCTCAGGTACTGATGAACACCTGTGTGTTGTTCACTTTCCCTC 43573
 |||||

QY 2262 ataaaaattgtcagctctctctgacacttagacctcaaaacttttagcatctctgtgagct 2321
 |||||

Db 43574 ATAAAAATTGTGAGCTCTCTCTGACACTTTAGACCTCAAACTTTAGCATCTCTGTGAGCT 43633
 |||||

QY 2322 gccatccactgtataatttcgctgctggaactggagtgagtggtgccccagggcagctg 2381
 |||||

Db 43634 GGCATCCACTGTATAATTTCGGCTGGCAACTGAGCTGAGGGAGTGTGCCAGGACGCTG 43693
 |||||

QY 2382 ccaagcaactccctcctggcttcagggtcagagtgcccaagcgttttatcagagcaacatc 2441
 |||||

Db 43694 CAAAGCACCTCCCTCCCTGGCTTCAGGGTCAGAGTGCCCGAGCGTTTATCAGAGGACGATC 43753
 |||||

QY 2442 caagcccagagcagtgctgactcttcggctggtgcttctcctctgaggggctatcaatg 2501
 |||||

Db 43754 CAAGCCACAGGCAGCTGTCGACTTTCGGCTGGTGGCTTTCTCTGAGGGGCTATCAATG 43813
 |||||

QY 2502 tgtagataaagccctgagtaggcaagagcagtgagatccactgctatgcttctgtatcat 2561
 |||||

Db 43814 TGTAGATAAAGCCCTGAGTAGGCAAGAGCAGTGAGATCCACTGCTATGCTTGTATACAT 43873
 |||||

QY 2562 cctcaaaccttccctcccccagcacagagaataattgctggcagcatcaaacctgcaaaagaa 2621
 |||||

Db 43874 CTTCAAACTTTCCCTTCCCGAGCAGAGGAATATTGGCTGGCATGCAACTGCAAAAGAA 43933
 |||||

QY 2622 aaatgcgaagcgccgggcacggtggtcctatgctctgaatcccccagcactttgggggctg 2681
 |||||

Db 43934 AAATGGGAAGCGCCGGGCACGGTGGCTCATGCTGTATATCCAGCAGCTTTGGGGGGCTG 43993
 |||||

QY 2682 aggtgggcgaatcatgagatcagggaggttcgagaccagcctggccagcatggtgaaacccc 2741
 |||||

Db 43994 AGGTGGCGCAATCATGAGATCAGGAGTTGCGAGACCAGCCTGGCCAGCATGGTGAACCCC 44053
 |||||

QY 2742 atctctactaaaaatacaaaaaattagctggcggtgtgacggcgccctgtaatccaca 2801
 |||||

Db 44054 ATCTCTACTAAAAAATAACAAAAATTAGCTGGCGGTGGTGACGGGCGCTGTAAATCCAGA 44113
 |||||

QY 2802 tactcaggagctgaggtaggagaaacactgaacctggaggtggaagtgcagtgagac 2861
 |||||

Db 44114 TACTCAGGAGCTGAGGTAGGAGAACTTGAACCTGGAGGTGGAAGTTGCAGTGAAC 44173
 |||||

QY 2862 caagatcacgcacactgcactccagcctggcgagtgagcgagactcccaactcaaaaaaa 2921
 |||||

Db 44174 CAAGATCACGCCACTGCATCCAGCCTGGCGCATGGAGCGAGACTCCAACTCAAAAAAAA 44233
 |||||

QY 2922 aaaaaaaa 2929
 |||||

Db 44234 AAAAGAA 44241
 |||||

RESULT 11

HSJ397H23/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Human DNA sequence from clone RP3-397H23 on chromosome 6q12-14.1,
 complete sequence.

AL121972

AL121972.17 GI:13276603

HTG

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 113202)

Direct Submission

Submitted (17-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Mar 12, 2001 this sequence version replaced gi:12832031.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>
 RP3-397H23 is from the library RP3-3 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pCYPAC2
 This sequence is the entire insert of clone RP3-397H23.

FEATURES

| source | Location/Qualifiers |
|---------------|--|
| | 1..113202 |
| | /organism="Homo sapiens" |
| | /db_xref="taxon:9606" |
| | /chromosome="6" |
| | /map="q12-14.1" |
| | /clone="RP3-397H23" |
| | /clone_lib="RP3-3" |
| | 1..58 |
| repeat_region | /note="AluJ/FLAM repeat: matches 5..62 of consensus" |
| repeat_region | 260..570 |
| repeat_region | /note="AluSx repeat: matches 1..305 of consensus" |
| repeat_region | 674..707 |
| repeat_region | /note="MER20 repeat: matches 29..62 of consensus" |
| repeat_region | 708..880 |
| repeat_region | /note="AluSp repeat: matches 120..292 of consensus" |
| repeat_region | 961..1095 |
| repeat_region | /note="AluSg1 repeat: matches 1..135 of consensus" |
| repeat_region | 1105..1428 |
| repeat_region | /note="AluSx repeat: matches 1..306 of consensus" |
| repeat_region | 1432..1585 |
| repeat_region | /note="MER20 repeat: matches 60..218 of consensus" |
| repeat_region | 1826..2123 |
| misc_feature | /note="AluY repeat: matches 1..295 of consensus" |
| misc_feature | complement(2112..2118) |
| misc_feature | /note="Single clone region. assembly confirmed by restriction digest data" |
| misc_feature | complement(2120..2127) |
| misc_feature | /note="Single clone region. assembly confirmed by restriction digest data" |
| misc_feature | complement(2129..2147) |
| misc_feature | /note="Single clone region. assembly confirmed by restriction digest data" |
| misc_feature | complement(2149..2160) |
| misc_feature | /note="Single clone region. assembly confirmed by restriction digest data" |
| repeat_region | 2150..2458 |
| misc_feature | /note="AluSx repeat: matches 1..306 of consensus" |
| misc_feature | complement(2161..2164) |
| misc_feature | /note="Single clone region. assembly confirmed by restriction digest data" |
| misc_feature | complement(2168..2263) |
| misc_feature | /note="Single clone region. assembly confirmed by restriction digest data" |
| repeat_region | 2468..2774 |
| repeat_region | /note="AluSg1 repeat: matches 1..308 of consensus" |
| repeat_region | 2795..2912 |
| repeat_region | /note="FLAM_C repeat: matches 1..118 of consensus" |
| repeat_region | 3028..3325 |
| repeat_region | /note="AluY repeat: matches 1..296 of consensus" |
| repeat_region | 3355..3644 |
| repeat_region | /note="AluSg repeat: matches 1..287 of consensus" |
| repeat_region | 3775..4074 |
| repeat_region | /note="AluSx repeat: matches 1..300 of consensus" |
| repeat_region | 4096..4386 |
| repeat_region | /note="AluJo repeat: matches 1..298 of consensus" |
| repeat_region | 4490..4793 |
| repeat_region | /note="AluJb repeat: matches 1..306 of consensus" |
| repeat_region | 5117..5253 |
| repeat_region | /note="FLAM_C repeat: matches 1..132 of consensus" |
| repeat_region | 5501..5814 |
| repeat_region | /note="AluSg repeat: matches 1..308 of consensus" |
| repeat_region | 6097..6225 |
| repeat_region | /note="FLAM_A repeat: matches 1..131 of consensus" |
| repeat_region | 6296..6479 |
| repeat_region | /note="AluJb repeat: matches 129..312 of consensus" |
| repeat_region | 6545..6830 |
| repeat_region | /note="AluJo repeat: matches 42..306 of consensus" |
| repeat_region | 6965..7270 |
| repeat_region | /note="AluSg repeat: matches 1..306 of consensus" |
| repeat_region | 7308..7602 |
| repeat_region | /note="AluJb repeat: matches 1..295 of consensus" |
| repeat_region | 7737..8035 |
| repeat_region | /note="AluSg repeat: matches 1..299 of consensus" |
| repeat_region | 8184..8488 |
| repeat_region | /note="AluSx repeat: matches 1..299 of consensus" |
| repeat_region | 8917..9176 |
| repeat_region | /note="AluSg1 repeat: matches 1..293 of consensus" |
| repeat_region | 9212..9377 |
| repeat_region | /note="L1MB4 repeat: matches 6005..6177 of consensus" |
| repeat_region | 9490..9785 |
| repeat_region | /note="AluSx repeat: matches 1..296 of consensus" |
| repeat_region | 10352..10431 |
| repeat_region | /note="L2 repeat: matches 2616..2702 of consensus" |
| repeat_region | 10432..10734 |
| repeat_region | /note="AluSx repeat: matches 1..301 of consensus" |
| repeat_region | 10735..10778 |
| repeat_region | /note="L2 repeat: matches 2575..2616 of consensus" |
| repeat_region | 10846..10985 |
| repeat_region | /note="MIR repeat: matches 53..200 of consensus" |
| repeat_region | 11023..11140 |
| repeat_region | /note="FLAM_C repeat: matches 1..118 of consensus" |
| repeat_region | 12005..12079 |
| repeat_region | /note="L1MD3 repeat: matches 7662..7737 of consensus" |
| repeat_region | 12350..12647 |
| repeat_region | /note="AluSx repeat: matches 1..297 of consensus" |
| repeat_region | 12669..12916 |
| repeat_region | /note="AluSx repeat: matches 1..293 of consensus" |
| repeat_region | 13058..13351 |
| repeat_region | /note="AluSc repeat: matches 1..296 of consensus" |
| repeat_region | 13504..13657 |
| repeat_region | /note="AluY repeat: matches 152..306 of consensus" |
| repeat_region | 13838..13947 |
| repeat_region | /note="AluSx repeat: matches 25..138 of consensus" |
| repeat_region | 13948..14254 |
| repeat_region | /note="AluSg repeat: matches 1..305 of consensus" |
| repeat_region | 14255..14437 |
| repeat_region | /note="AluSx repeat: matches 138..307 of consensus" |
| repeat_region | 14482..14610 |
| repeat_region | /note="MERSA repeat: matches 46..167 of consensus" |
| repeat_region | 14614..14911 |
| repeat_region | /note="AluSc repeat: matches 1..301 of consensus" |
| repeat_region | 14934..15229 |
| repeat_region | /note="AluSp repeat: matches 1..297 of consensus" |
| repeat_region | 15266..15556 |
| repeat_region | /note="AluJo repeat: matches 5..299 of consensus" |
| repeat_region | 15578..15888 |
| repeat_region | /note="AluYa5 repeat: matches 1..311 of consensus" |
| repeat_region | 15965..16082 |
| repeat_region | /note="FLAM_C repeat: matches 1..118 of consensus" |
| repeat_region | 16157..16439 |
| repeat_region | /note="AluJb repeat: matches 11..294 of consensus" |
| repeat_region | 16618..16913 |
| repeat_region | /note="AluSp repeat: matches 1..296 of consensus" |

repeat_region 16933..16998
/note="33 copies 2 mer tt 69% conserved"
repeat_region 17001..17104
/note="AluSg/x repeat: matches 196. .307 of consensus"
repeat_region 17135..17428
/note="AluSg repeat: matches 1. .293 of consensus"
repeat_region 17790..18090
/note="AluSx repeat: matches 1. .301 of consensus"
repeat_region 19220..19547
/note="AluB repeat: matches 2. .290 of consensus"
repeat_region 20724..21025
/note="AluJo repeat: matches 1. .304 of consensus"
repeat_region 21045..21082
/note="19 copies 2 mer gt 89% conserved"
repeat_region 21084..21260
/note="AluSg/x repeat: matches 134. .310 of consensus"
repeat_region 21367..21670
/note="AluSx repeat: matches 1. .306 of consensus"
repeat_region 21819..21939
/note="MIR repeat: matches 32. .165 of consensus"
repeat_region 22647..22943
/note="AluSx repeat: matches 1. .299 of consensus"
repeat_region 23011..23316
/note="AluY repeat: matches 1. .306 of consensus"
repeat_region 23399..23702
/note="AluSg repeat: matches 1. .304 of consensus"
repeat_region 24322..24622
/note="AluSx repeat: matches 1. .302 of consensus"
repeat_region 24680..24971
/note="AluSc repeat: matches 1. .292 of consensus"
repeat_region 25178..25302
/note="AluSg/x repeat: matches 172. .300 of consensus"

Query Match 40.7%; Score 1193.6; DB 9; Length 113202;
Best Local Similarity 99.3%; Pred. No. 3.5e-271; Mismatches 9; Indels 0; Gaps 0;
Matches 1199; Conservative 0;

QY 1722 gtctgacctgataacactgttgagaatggcaaacctgttctctattgtgtgctgcta 1781
|||||
Db 56637 GTCTCTCTTTATAGAACACTGTGTGGAGATGGCAACCGTGTCTATATTGCTGCTGCTA 56578

QY 1782 ttaatgttttggcgccattttttacactattcccaaaagtgaaagtcacaaactggg 1841
|||||
Db 56577 TTAATGTTTTTGGTGCATTTCTTTTACACTATTTCGCCAAAGGTAAGTACAAAACGTGG 56518

QY 1842 cctcaatgatacactgacacacactgaagaaacaaataataatcctgcctctat 1901
|||||
Db 56517 CTCTCAATGATCACCATGACACAGACTGAAGGAACCAATAATAATCCTGCCTCTAT 56458

QY 1902 taatgtatttttatttatcatgttaacctcaaaagtccttctgtattgtgtaagcattcta 1961
|||||
Db 56457 TAATGTATTTTATTTATCATGTAACTCAAGTGCCTTCTGTATTGTGTAAAGCATCTTA 56398

QY 1962 tgtcttttttaattgtaactgtattagatttttaaggccataatcatgaataatcaact 2021
|||||
Db 56397 TGTCTTTTTTTAATGCTACTGTATTAGATTTTAAAGGCTATAAATCATGAATATCACT 56338

QY 2022 agttgccagaataataaagtaactgttttaattatgaataatgaactaggacit 2081
|||||
Db 56337 AGTTGCCAGAAATAAATGAATGAACTGTGTTAATTATGAATAATATGTAAAGCTAGGACTT 56278

QY 2082 ctacttttagttcacatacctgctgctagtcgggcaacatgaagtaggacatttctgtt 2141
|||||
Db 56277 CTACTTTTAGTTCACATACCTGCCTGCTAGTCGGCAACATGAAGTAGGACACTTCTGTT 56218

QY 2142 gatttttgagccataactaaaggggaatgagctgaacagacctcctgtatcactttgctt 2201
|||||
Db 56217 GATTTTTTAGGGCCACTAAAGGGAATGAGCTGAAACAGACACCTCTCTGATACCTTTGCTT 56158

QY 2202 aattaactagatgataaattctcagttactgataataaacacctgttgtttcactttcctc 2261
|||||
Db 56157 AATTAACCTAGATGATAATTCTCAGTACTGATAAACACACTGTTGTTGTTCACTTTCCCTC 56098

QY 2262 ataaaattgtcagctctctctgacacttagacactcaaaacttttagcatctctgtgagct 2321
|||||
Db 56097 ATAAAAATTTGTCAGCTCTCTCTGACACTTAGACCTCAAACTTTAGCATCTCTCTGAGCT 56038

QY 2322 gccatccactgtataatttcgctctgcaactggaactgagggagtggtgccaggaagctg 2381
|||||
Db 56037 GCCATCCACTGTATAATTTTCGCTTGGCAACTGAGCTGAGGGAGTGTGCCAGGCAGCTG 55978

QY 2382 caaagcactccctccctgcttcaggttcagagtgcccagcgtttatcagagagcagcatc 2441
|||||
Db 55977 CCAAGCAGCTCCCTCCCTGCTTCAGGGTCAGAGTGCCAGCGTTTATCAGAGCAGCATC 55918

QY 2442 caagccagagccagtgctgactcttcggctgtggtccttctctctgaggggcttatcaatg 2501
|||||
Db 55917 CAAGCCAGAGCCAGTGTGCACTCTTCGGCTGGTGCTTTCTCTCAGGGGCTATCAATG 55858

QY 2502 ttagataaagccctgagtaggcaagagcagtgagatccactgctatggtcttctgtacatat 2561
|||||
Db 55857 TGTAGATAAAGCCCTGAGTAGGCAAGAGCAGTGAATCCACTGCTATGTCTTTGATACAT 55798

QY 2562 cctcaaaacttccctcccgacacagagaatattgctggtgcatgcaacctgcaaaagaa 2621
|||||
Db 55797 CCTCAAACTTTCCTTCCAGCAGACAGAGGAATATTGGCTGGCATGCAACTGCAGAAAGAA 55738

QY 2622 aaatgcgaagcgccgggcacggtgctcatgctgttaatcccgacactttgggggctg 2681
|||||
Db 55737 AAATGCGAAGCGCGCGGACGGTGCTCATGCTCTAATCCAGCACCTTTGGGGGGCTG 55678

QY 2682 agtgggcgaatcatgagatcaggaggttcgagaccagcctggccagcatggtaaacccc 2741
|||||
Db 55677 AGGTGGGGCAATCATGAGATCAGGAGTTCGAGACCAGCCTGGCCAGCATGGTGAACCCC 55618

QY 2742 atctctactaaaaatacaaaaaattagctggcgctggtgacggcgccctgttaatccaga 2801
|||||
Db 55617 ATCTCTACTATAAATAACAAAAAATTAGCTGGCGCTGGTGACGGGCCCTGTATATCCACA 55558

QY 2802 tactcaggaggtgaggtgaggagaatcactgaaacctgggaggtggaagtgcagtgaac 2861
|||||
Db 55557 TACTCAGGAGGCTGAGGTAGGAGAATCACTTGAACCTGGGAGGTGGAAGTTGCAGTGAAC 55498

QY 2862 caagatcaagccactcactccagcctggcgatggcagatgagcgagagactccaaactcaaaaaa 2921
|||||
Db 55497 CAAGATCACGCCACTCACTCCAGCCTGGCGATGGAGCAGAGACTCCAATCAAAANAAA 55438

QY 2922 aaaaaaaa 2929
Db 55437 AAAAAGAA 55430

RESULT 12
AC034271 149597 bp DNA linear HTG 12-MAY-2000
LOCUS Homo sapiens chromosome 6 clone RP11-598G19 map 6, WORKING DRAFT
DEFINITION SEQUENCE, 31 unordered pieces.
ACCESSION AC034271
VERSION AC034271.2 GI:7770480
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 149597)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 6, clone RP11-598G19
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 149597)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boquslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodgse,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEvan, P., McGuck, A., McKernan, K., McPheeters, R., Melidrim, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (05-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 12, 2000 this sequence version replaced gi:7417823.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7098

Center clone name: 598_G_19

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 134701 bases at least Q40

Consensus quality: 141812 bases at least Q30

Consensus quality: 144823 bases at least Q20

Insert size: 166000; agarose-ff

Quality coverage: 3.6 in Q20 bases; sum-of-contigs

Quality coverage: 4.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 31 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 1018: contig of 1018 bp in length
* 1019 1118: gap of 100 bp
* 1119 2203: contig of 1085 bp in length
* 2204 2303: gap of 100 bp
* 2304 3814: contig of 1511 bp in length
* 3815 3914: gap of 100 bp
* 3915 4327: contig of 413 bp in length
* 4328 4427: gap of 100 bp
* 4428 6447: contig of 2020 bp in length
* 6448 6547: gap of 100 bp
* 6548 8884: contig of 2337 bp in length
* 8885 8984: gap of 100 bp
* 8985 10630: contig of 1646 bp in length
* 10631 10730: gap of 100 bp
* 10731 13203: contig of 2472 bp in length
* 13203 13302: gap of 100 bp
* 13303 15992: contig of 2690 bp in length
* 15993 16092: gap of 100 bp
* 16093 18722: contig of 2630 bp in length
* 18723 18822: gap of 100 bp
* 18823 21695: contig of 2873 bp in length
* 21696 21795: gap of 100 bp
* 21796 25265: contig of 3470 bp in length
* 25266 25365: gap of 100 bp

```

```

* 25366 28014: contig of 2649 bp in length
* 28015 28114: gap of 100 bp
* 28115 31378: contig of 3264 bp in length
* 31379 31478: gap of 100 bp
* 31479 34284: contig of 2806 bp in length
* 34285 34384: gap of 100 bp
* 34385 37707: contig of 3323 bp in length
* 37708 37807: gap of 100 bp
* 37808 41922: contig of 4115 bp in length
* 41923 42022: gap of 100 bp
* 42023 45878: contig of 3856 bp in length
* 45879 45978: gap of 100 bp
* 45979 49890: contig of 3912 bp in length
* 49891 49990: gap of 100 bp
* 49991 54041: contig of 4051 bp in length
* 54042 54141: gap of 100 bp
* 54142 59531: contig of 5390 bp in length
* 59532 59631: gap of 100 bp
* 59632 65537: contig of 5906 bp in length
* 65538 65637: gap of 100 bp
* 65638 71486: contig of 5849 bp in length
* 71487 71586: gap of 100 bp
* 71587 76694: contig of 5108 bp in length
* 76695 76794: gap of 100 bp
* 76795 82579: contig of 5785 bp in length
* 82580 82679: gap of 100 bp
* 82680 90920: contig of 8241 bp in length
* 90921 91020: gap of 100 bp
* 91021 99034: contig of 8014 bp in length
* 99035 99134: gap of 100 bp
* 99135 110324: contig of 11190 bp in length
* 110325 110424: gap of 100 bp
* 110425 119810: contig of 9386 bp in length
* 119811 119910: gap of 100 bp
* 119911 130142: contig of 10232 bp in length
* 130143 130242: gap of 100 bp
* 130243 149597: contig of 19355 bp in length.

```

Location/Qualifiers

source

```

1. .149597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6"
/clone="RP11-598G19"
/clone_lib="RPC1-11 Human Male BAC"
1. .1018
/note="assembly_fragment"
1119. .2203
/note="assembly_fragment"
2304. .3814
/note="assembly_fragment"
3915. .4327
/note="assembly_fragment"
clone_end:r7
vector_side:right
4428. .6447
/note="assembly_fragment"
6548. .8884
/note="assembly_fragment"
8985. .10630
/note="assembly_fragment"
10731. .13202
/note="assembly_fragment"
13303. .15992
/note="assembly_fragment"
16093. .18722
/note="assembly_fragment"
18823. .21695
/note="assembly_fragment"
21796. .25265
/note="assembly_fragment"
25366. .28014
/note="assembly_fragment"

```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

```
misc_feature 28115..31378
/note="assembly_fragment"
misc_feature 31479..34284
/note="assembly_fragment"
misc_feature 34385..37707
/note="assembly_fragment"
misc_feature 37808..41922
/note="assembly_fragment"
misc_feature 42023..45878
/note="assembly_fragment"
misc_feature 45979..49890
/note="assembly_fragment"
misc_feature 49991..54041
/note="assembly_fragment"
misc_feature 54142..59531
/note="assembly_fragment"
misc_feature 59632..65537
/note="assembly_fragment"
misc_feature 65638..71486
/note="assembly_fragment"
misc_feature 71587..76694
/note="assembly_fragment"
misc_feature 76795..82579
/note="assembly_fragment"
misc_feature 82680..90920
/note="assembly_fragment"
misc_feature 91021..99034
/note="assembly_fragment"
misc_feature 99135..110324
/note="assembly_fragment"
misc_feature 110425..119810
/note="assembly_fragment"

Query Match 40.7%; Score 1192; DB 2; Length 149597;
Best Local Similarity 99.2%; Pred. No. 8.9e-271;
Matches 1198; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1722 gtctgacccctgatacacactgttgagaaatggcaaacctgttctctattattgctgctgcta 1781
|||||
Db 75367 GTCTTCTTTATAGAACACTGTTGGAGATGGCAACCGTTCATATTGCTGCTGCTA 75426

QY 1782 ttaatgttttggcgcaattttcttacactattcgcaaaagtgaaagtcacaaactggg 1841
|||||
Db 75427 TTAATGTTTTTGGTGGCATTTCTTTACACTATTGCCAAAGGTGAAGTACAAACCTGGG 75486

QY 1842 ctctcaatgataccatgacacagacactgaaggaaccaataaataactcctcctctat 1901
|||||
Db 75487 CTCTCAATGATCACATGGACACAGACACTGAAGGAACCAATAAATAATCCTGCCCTCTAT 75546

QY 1902 taatgtatttttatcatgttaacctcaaaagtgccttctgtattgtgttaagcattcta 1961
|||||
Db 75547 TAATGTATTTTATTTATCATGTAACTCAAGTGCCTTCTGTATTGTGTAAAGCATTCCTA 75606

QY 1962 tgccttttttaattgtactgttattagatttttaaggccctataatgaataatacaact 2021
|||||
Db 75607 TGCCTTTTTTAAATGCTACTGTATTAGATTTTAAAGGCCCTATAAATCATGAAATATCACT 75666

QY 2022 agttgcagaataataaataaactgttttaattatgaataataatgaactagacatt 2081
|||||
Db 75667 AGTTGCCAGAAATATAAATGAACCTGTGTTAATTATGAATAATATGTAAGCTAGGACTT 75726

QY 2082 ctactttagttcacatacctgctgctagtcgagcgggcaacatgaagttaggacagttctgtt 2141
|||||
Db 75727 CTACTTTAGTTTCACATACCTGCTGCTAGTCGGGCAACATGAAGTAGGACACTTCTGTT 75786

QY 2142 gattttttgggccataactaaaggggaatgagctgaacagacctcctgtatcaccttgcctt 2201
|||||
Db 75787 GATTTTTTAGGGCCATCTAAAGGAATGAGCTGGAACAGACCTCCTGTATACCTTTGCTT 75846

QY 2202 aattaactagatgaataattctcagttactgataacacacctgttgtttgttcaacttccctc 2261
|||||
Db 75847 AATTAACATAGATGATAATTCTCAGGTACTGATAAACACCTGTTGTGTTCCTCTCCCTC 75906
```

```
QY 2262 ataaaattgtcagctctctctgacacttagaacctcaaaacttttagcatctctgtgagct 2321
|||||
Db 75907 ATAAAAATTGTCAGCTCTCTCTGACACTTAGACCTCAAACTTTAGCATCTCTGTGGAGCT 75966

QY 2322 gccatccactgtataatttcgacctggcaactggaactgaggggagtggtgccagagcagctg 2381
|||||
Db 75967 GCCATCCACTGTATAATTTCGCTGGCACTGAGGAGAGTGTCGCCAGGCAAGCTG 76026

QY 2382 ccaagcactctccctggcttcagaggtcagagtcgccagcgttttatcagagcagcattc 2441
|||||
Db 76027 CCAAGCACTCCCTCCCTGGCTTCAGGCTCAGAGTGCCAGCGTTTATCAGAGGACGATC 76086

QY 2442 caagccagagccagtgctgactcttcggctggctgttctctctgaggggcttatcaatg 2501
|||||
Db 76087 CAAGCCAGAGCCAGTGTGACTCTTCGGCTGCTGCTTTCCCTCTGAGGGGCTATCAATG 76146

QY 2502 tgtagataaagccctgagtaggcaagcagtgagatccaactgctatggtctctgatacat 2561
|||||
Db 76147 TGTAGATAAAGCCCTGAGTAGGCAAGCAGCAGTAGATCCACTGCTATGGTCTTTGATACAT 76206

QY 2562 cctcaaaacttctctccagcacagagaatattggctggcatgcaacctgcaaaagaa 2621
|||||
Db 76207 CCTCAAACTTTCCCTCCCAAGCAGAGGAAATATGGCTGGCATGCAACTGCAAAAGAA 76266

QY 2622 aaatgcgaagcggcgacggtggtcctatgctctgtaatcccagcactttgggggctg 2681
|||||
Db 76267 AAATGCGAAGCGCGGCGGACGCTGCTCATGCTGTATATCCCAAGCAGCATTGTTGGGGCTG 76326

QY 2682 agtgggcgaaatcatgatcagtaggttcagaccagcctggccagcatggtgaaacccc 2741
|||||
Db 76327 AGTGGGCGAATCATGAGATCAGAGAGTTTCGAGACCAAGCTGGCCAGCATGGTGAACCCCC 76386

QY 2742 atcttactaaaaatacaaaaaattagctggcgctggtgaacggcgccctgtaatccacaga 2801
|||||
Db 76387 ATCTTACTAAAAATACAAAAAATTAGCTGGCGCTGCTGAGCGGCGCTGTAAATCCAGAG 76446

QY 2802 tactcaggaagctgagtagagaatcacctggaacctggaggtggaggttcagtgaaac 2861
|||||
Db 76447 TACTCAGGAGGCTGAGTAGGAGAATCACATTGAACCTGGGAGATGGAAGTTGCAGTGAAC 76506

QY 2862 caagatacagccactgcactccagcctggcgatgagcagagactcccaactcaaaaaaa 2921
|||||
Db 76507 CAAGATCACCCCACTGCCTCCAGCTGGCGCATGGCGAGAGACTCCAACTCAAAAAAAA 76566

QY 2922 aaaaaaaa 2929
|||||
Db 76567 AAAAGAA 76574
```

```
RESULT 13
AC025535/c 157749 bp DNA linear HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 6 clone RP11-206H23, WORKING DRAFT
DEFINITION SEQUENCE, 25 unordered pieces.
ACCESSION AC025535
VERSION AC025535.4 GI:7658462
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 157749)
JOURNAL Waterston, R.H.
AUTHORS The sequence of Homo sapiens clone
TITLE Unpublished
JOURNAL 2 (bases 1 to 157749)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) Genome Sequencing Center, Washington
MO 63108, USA
COMMENT On Apr 28, 2000 this sequence version replaced gi:7582731.
```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0206H23
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator ET; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 141377 bases at least Q40
Consensus quality: 146682 bases at least Q30
Consensus quality: 148842 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 155349; sum-of-contigs
Quality coverage: 3.52 in Q20 bases; agarose-fp
Quality coverage: 3.92 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1470: contig of 1470 bp in length
* 1471 1570: gap of unknown length
* 1571 3019: contig of 1449 bp in length
* 3020 3119: gap of unknown length
* 3120 5356: contig of 2237 bp in length
* 5357 5456: gap of unknown length
* 5457 7747: contig of 2291 bp in length
* 7748 7847: gap of unknown length
* 7848 9214: contig of 1366 bp in length
* 9214 9313: gap of unknown length
* 9314 10753: contig of 1440 bp in length
* 10754 10853: gap of unknown length
* 10854 12243: contig of 1390 bp in length
* 12244 12343: gap of unknown length
* 12344 14551: contig of 2208 bp in length
* 14552 14651: gap of unknown length
* 14652 17066: contig of 2415 bp in length
* 17067 17166: gap of unknown length
* 17167 19602: contig of 2436 bp in length
* 19603 19702: gap of unknown length
* 19703 23843: contig of 4141 bp in length
* 23844 23943: gap of unknown length
* 23944 27783: contig of 3840 bp in length
* 27784 27883: gap of unknown length
* 27884 31792: contig of 3909 bp in length
* 31793 31892: gap of unknown length
* 31893 35687: contig of 3795 bp in length
* 35688 35787: gap of unknown length
* 35788 39711: contig of 3924 bp in length
* 39712 39811: gap of unknown length
* 39812 44204: contig of 4393 bp in length
* 44205 44304: gap of unknown length
* 44305 48833: contig of 4529 bp in length
* 48834 48933: gap of unknown length
* 48934 56617: contig of 7684 bp in length
* 56618 56717: gap of unknown length
* 56718 63063: contig of 6346 bp in length
* 63064 63163: gap of unknown length
* 63164 74288: contig of 11125 bp in length
* 74289 74388: gap of unknown length
* 74389 87403: contig of 13015 bp in length
* 87404 87503: gap of unknown length
* 87504 99698: contig of 12195 bp in length
* 99699 99798: gap of unknown length
* 99799 115384: contig of 15586 bp in length

* 115385 115484: gap of unknown length
* 115485 132131: contig of 16647 bp in length
* 132132 132231: gap of unknown length
* 132232 157749: contig of 25518 bp in length.
FEATURES
source location/Qualifiers
1. .157749
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-206H23"
BASE COUNT 43129 a 34263 c 34620 g 43330 t 2407 others
ORIGIN
Query Match 40.7%; Score 1192; DB 2; Length 157749;
Best Local Similarity 99.2%; Pred. No. 9e-271;
Matches 1198; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
QY 1722 gctcgaccctgataaacactgttgagaaatggcaaacccgtgttctatatattgctgctgcta 1781
||||| |
Db 93294 GTCCTCCTTTATAGAACACACTGTTGGAGATGGCAACCGTGTCTATATTGCTGCTGCTA 93235
QY 1782 ttaatgttttggcgcaatttctttacactattcgcaaaagtgcaagtcacaaaactg99 1841
||||| |
Db 93234 TTAATGTTTTGGTGCCATTTCTTTACACTATTGCCAAAGCTGAAGTACAAAACCTGGG 93175
QY 1842 ctctcaatgacacacagacacagacactgaaggaacaaataaataaactcctccttat 1901
||||| |
Db 93174 CTCCTAATGATCACCATGGACACAGACACTGAAGGAACCAATAAATAATCCTGCCCTCTAT 93115
QY 1902 taatgtatttttatttatcatgtaacctaaagtcgctctctgtattgtgtaagcattcta 1961
||||| |
Db 93114 TAATGTATTTTATTATCATGTAACTCAAAAGTGCTTCTGTATTGTGTAAACATCTCA 93055
QY 1962 tgccttttttaattgactgtattagatttttaaggccataataatgaataatacact 2021
||||| |
Db 93054 TGTCTTTTTTAAATGCTACTGTATTAGATTTTAAAGGCCATATAAATCATGAATAATCACT 92995
QY 2022 agttgcagaataaataaataaactgttttaattatgaataatgaataatgaactagactt 2081
||||| |
Db 92994 AGTTGCCAGAAATATAAATGAACACTGTGTATTATTAATTAATAATATGTAAGCTAGGACTT 92935
QY 2082 ctactttaggttcaacactgctgctgtagtggcggaacatgaagttaggacagtctgtt 2141
||||| |
Db 92934 CTACTTTAGTTTCACATACCTGCTGCTAGTGGGCAACATGAAGTAGGACAGCTTCTGTT 92875
QY 2142 gatttttagggccataactaaaggggaatgagctgaacagacctcctgatacctttgctt 2201
||||| |
Db 92874 GATTTTTTAGGGCCATACTAAAGGGAATGAGCTGAACACAGACCTCTGTATACCTTTGCTT 92815
QY 2202 aattaactagatgataattctcaggtactgataaacacctgtgtttcactttcctc 2261
||||| |
Db 92814 AATTAAACTAGATGATAATTCTCAGGTACTGATAAACACCTGTTGTGTTCATTCTTCCCTC 92755
QY 2262 aaaaaattgtcagctctctctgacacttagacctcaaaacttttagcatctctgtggagct 2321
||||| |
Db 92754 ATAAAAATTGTCAGCTCTCTCTGACACTTAGACCTCAAACTTTAGCATCTCTGTGAGCT 92695
QY 2322 gccatccactgtataatttcgcctggcgaactggagtgaggagtgccccagggcagctg 2381
||||| |
Db 92694 GCCATCCACTGTATAAATTTCCGCTGGCAACTGGACTGAGGGGAGTGTGCCAGGCAGCTG 92635
QY 2382 ccaagcaactccctcctggcttcagggtcagagtgcccaagcgttttatacagaggaacalc 2441
||||| |
Db 92634 CCAAGCACCTCCCTCCCTGGCTTCAGGGTCAGAGTGGCCAGCGTTTATCAGAGGACGATC 92575
QY 2442 caagccagagccagtgctgactcttcggctggtggtccttctcctctgaggggctatacaatg 2501
||||| |
Db 92574 CAAGCCACAGGCAGGTGTCGACTCTTCGGCTGGTGGCTTTCTCTCAGGGGCTATCAATG 92515
QY 2502 tgtagataaagccctgagtaggcaagagcagtgagataccactgctatggtctcttgatacat 2561
||||| |
Db 92514 TGTAGATAAAGCCCTGTAGTAGGCAGAGCAGTGAATCCTGCTATGTTGTTGATACAT 92455

QY 2562 cctcaacttccctccagcagaggaattattgctggcatgcaacctgcaaaagaa 2621
|||||
Db 92454 CCTCAACTTTCCTCCAGCAGACAGAGGATATTGGCTGGCATGCAACTGCAAAAGAA 92395
|||||
QY 2622 aaatgcgaagcgccggcagcagtgctcctgctgaatccccagcacttgggggctg 2681
|||||
Db 92394 AAATGGGAAGCGCGGCGACGCTGCTATGCTCTAATATCCAGCAGCATTGGGGGGCTG 92335
|||||
QY 2682 agtgggcaaatcatagatcaagagttcgagacagcctggccagcatgtgtaaacccc 2741
|||||
Db 92334 AGTGGCGCAANTCATGAGATCAGAGTTCGAGACCAAGCCTGGCCAGCATGGTGAAACCCC 92275
|||||
QY 2742 atctctactaaaaatacaaaaaattagctggcgctggtgacggcgccctgtaattccaga 2801
|||||
Db 92274 ATCTCTACTAAAAATACAAAAATTAGCTGGCGCTGTGACGGCGCCTGTATATCCAGA 92215
|||||
QY 2802 tactcagggagctgaggtaggagaatacaacttgaaactggaggtggaagtgcagtgaac 2861
|||||
Db 92214 TACTCAGGAGGCTGAGGTAGGAGAAATCACTGAACCTGGGAGATGGAAGTTGCAGTGAAC 92155
|||||
QY 2862 caagatcacgcactgcactccagcctggcgctgagcgagactccaactcaaaaaaa 2921
|||||
Db 92154 CAGATCACGCCACTGCATCTCCAGCTGGCGATGGAGCGAGAGACTCAACTCAAAAAAAA 92095
|||||
QY 2922 aaaaaaa 2929
|||||
Db 92094 AAAAAGAA 92087
|||||
RESULT 14
AL590428/c
LOCUS
DEFINITION Human DNA sequence from clone Rp11-553A21 on chromosome 6, complete
sequence.
ACCESSION AL590428 AC028605
VERSION AL590428.7 GI:15072593
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 163577)
Direct Submission
Submitted (31-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Aug 1, 2001 this sequence version replaced gi:15021177.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
Rp11-553A21 is from the library RPC1-11.2 constructed by the group
of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone
Rp11-553A21. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.

The true right end of clone Rp11-553A21 is at 163577 in this
sequence. The true left end of clone Rp11-525G3 is at 88067 in this
sequence. The true right end of clone Rp3-397H23 is at 2000 in this
sequence.

FEATURES

| Source | Location/Qualifiers |
|---------------|---|
| repeat_region | 1..163577 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="6" /clone="Rp11-553A21" /clone_lib="RPC1-11.2" 285..433 |
| repeat_region | /note="AluJb repeat: matches 1..119 of consensus" |
| repeat_region | 434..743 /note="AluY repeat: matches 1..311 of consensus" |
| repeat_region | 744..919 /note="AluJb repeat: matches 119..305 of consensus" |
| repeat_region | 988..1075 /note="L2 repeat: matches 2626..2708 of consensus" |
| repeat_region | 1166..1384 /note="L1MC5 repeat: matches 7317..7536 of consensus" |
| repeat_region | 1385..1690 /note="AluX repeat: matches 1..306 of consensus" |
| repeat_region | 1691..1759 /note="L1MC5 repeat: matches 7536..7604 of consensus" |
| repeat_region | 1760..2024 /note="AluY repeat: matches 35..287 of consensus" |
| repeat_region | 2025..2248 /note="L1MC5 repeat: matches 7604..7887 of consensus" |
| repeat_region | 2588..2706 /note="L1MC/D repeat: matches 5389..5514 of consensus" |
| misc_feature | 3686..4489 /note="CpG island" /evidence=not_experimental |
| repeat_region | 3695..3736 /note="21 copies 2 mer cc 78% conserved" |
| repeat_region | 4691..4996 /note="AluJo repeat: matches 3..306 of consensus" |
| repeat_region | 5628..5977 /note="MLT2PB repeat: matches 1..403 of consensus" |
| repeat_region | 5978..6030 /note="MER5A repeat: matches 44..97 of consensus" |
| repeat_region | 6118..6474 /note="L1MC3 repeat: matches 7398..7739 of consensus" |
| repeat_region | 6475..6771 /note="AluY repeat: matches 1..290 of consensus" |
| repeat_region | 6772..7373 /note="L1MC3 repeat: matches 6740..7398 of consensus" |
| repeat_region | 7491..7764 /note="AluJo repeat: matches 1..286 of consensus" |
| repeat_region | 7887..8168 /note="AluSg repeat: matches 1..287 of consensus" |
| repeat_region | 8365..8430 /note="33 copies 2 mer tc 89% conserved" |
| repeat_region | 8431..8729 /note="AluSg repeat: matches 1..302 of consensus" |
| repeat_region | 8779..8975 /note="MER33 repeat: matches 129..323 of consensus" |
| repeat_region | 8976..9292 /note="AluX repeat: matches 1..311 of consensus" |
| repeat_region | 9293..9415 /note="MER33 repeat: matches 1..129 of consensus" |
| repeat_region | 9516..9647 /note="MIR repeat: matches 106..226 of consensus" |
| repeat_region | 9698..9994 /note="AluSg repeat: matches 1..295 of consensus" |
| repeat_region | 10005..10048 /note="22 copies 2 mer tt 79% conserved" |

TITLE
Homo sapiens chromosome 6, clone RP11-598G19
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 149597)
AUTHORS
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bada,F.,
Boguslavskiy,L., Boukhgaiter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE
Direct Submission
JOURNAL
Submitted (05-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT
On May 12, 2000 this sequence version replaced gi:7417823.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7098
Center clone name: 598_G_19
----- Summary Statistics
Sequencing vector: M13; W7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 134701 bases at least Q40
Consensus quality: 141812 bases at least Q40
Consensus quality: 14823 bases at least Q30
Insert size: 166000; agarose-fp
Insert size: 146597; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 4.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* i
* 1019 1118: contig of 1018 bp in length
* 1119 2203: contig of 1085 bp in length
* 2204 2303: gap of 100 bp
* 2304 3814: contig of 1511 bp in length
* 3815 3914: gap of 100 bp
* 3915 4327: contig of 413 bp in length
* 4328 4427: gap of 100 bp
* 4428 6447: contig of 2020 bp in length
* 6448 6547: gap of 100 bp
* 6548 8884: contig of 2337 bp in length
* 8885 8984: gap of 100 bp
* 8985 10630: contig of 1646 bp in length
* 10631 10730: gap of 100 bp

```
misc_feature 13303..15992
/note="assembly_fragment"
misc_feature 16093..18722
/note="assembly_fragment"
misc_feature 18823..21695
/note="assembly_fragment"
misc_feature 21796..25265
/note="assembly_fragment"
misc_feature 25366..28014
/note="assembly_fragment"
misc_feature 28115..31378
/note="assembly_fragment"
misc_feature 31479..34284
/note="assembly_fragment"
misc_feature 34385..37707
/note="assembly_fragment"
misc_feature 37808..41922
/note="assembly_fragment"
misc_feature 42023..45878
/note="assembly_fragment"
misc_feature 45979..49890
/note="assembly_fragment"
misc_feature 49991..54041
/note="assembly_fragment"
misc_feature 54142..59531
/note="assembly_fragment"
misc_feature 59632..65537
/note="assembly_fragment"
misc_feature 65638..71486
/note="assembly_fragment"
misc_feature 71587..76694
/note="assembly_fragment"
misc_feature 76795..82579
/note="assembly_fragment"
misc_feature 82680..90920
/note="assembly_fragment"
misc_feature 91021..99034
/note="assembly_fragment"
misc_feature 99135..110324
/note="assembly_fragment"
misc_feature 110425..119810
/note="assembly_fragment"
```

```
Query Match 13.5%; Score 396.8; DB 2; Length 149597;
Best Local Similarity 87.1%; Pred. No. 6.9e-83;
Matches 474; Conservative 0; Mismatches 7; Indels 63; Gaps 1;

QY 1 gttcgtcgaagccctcccttaattatgtgcaattcaagtccccactgcccgcgcgcaa 60
|||||
Db 102673 GTTCGGTCCAAGCCCTCCCTTAATTATGTGCAATTCAAGTCCCCACTGCCCGCCGCAA 102614

QY 61 gccccactcactcgtcgtcggcgaggtggccctgcactttacaaggggtgcagga 120
|||||
Db 102613 GCCCCACATCATCTCGCTCGGGCAGGGTGGCCCTGCACATTTACAAGGGGTGCAGGA 102554

QY 121 gcgggagaggtgtctccgaacacggctcccggcgatgtagaccgagcgagcgagcg 180
|||||
Db 102553 GCGGGAGACGGTCGTCGGAACACGGCTCCCGGATGGTTAAACCGGGCGGGAGCGG 102494

QY 181 gctcaatttgcgccaatcctacagagaactcccagaactccgttccctagtccaacccaa 240
|||||
Db 102493 GCTCACTTTGCGCCAATCTCTACGAGAATCCCAAGAACTCCGCTTCCTAGTCCAACCAA 102434

QY 241 gccagattgccacacctaagatgagcgggggcg----- 277
|||||
Db 102433 GCCAGAGTTGCCCCACACCTAAATAATGGCGGGGGGAGTGGCGCGCGCCCTCTGGG 102374

QY 278 -----atgacacggccccgcgcggt 297
|||||
Db 102373 CGGGACCGGGGACTAGACGTGGCGGGGGGGTGTCTATCGCCCCCGCCCGCCGGT 102314

QY 298 ccagccagctcgcccggttcgggtgtctggtggcggtccctctctctgcaggt 357
|||||
```

```
Db 102313 CCAGCCAGCTCGGCCGGGGGCTTCGGGCTGTGGGCGCGCGCTCCCTTCTCTCCAGGT 102254
QY 358 ggcaggtacacctgtctcacgtagcggtcatgaggtcttcgggttcgagacctggcccgaa 417
|||||
Db 102253 GCGAGTACACCTGCTCAGTAGGCGTCAATGAGTCTCCGCTTCGAGACTGCGCGGAA 102194
|||||
QY 418 cgtatggcgagagacacgagaccgacgctcttaccgggagcgccacgagcgagc 477
|||||
Db 102193 CGATGGCGAGGAGACGAGACCGCACGCGCTCTTCTTACCGGGGCGCCCGGCGGAGC 102134
QY 478 cgct 481
|||||
Db 102133 CGGT 102130
```

Search completed: July 15, 2002, 13:56:39
Job time: 15364 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 08:24:01 ; Search time 56.81 Seconds
(without alignments)
967.814 Million cell updates/sec

Title: US-09-776-865-4
Perfect score: 2617
Sequence: 1 MKSPVSLAPSDGEGSDRT.....LFAKGEVQNWAISHQGHNRN 495

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/hold-geneseq/genesecp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 2617 | 100.0 | 495 | 21 | AA1980 |
| 2 | 2617 | 100.0 | 495 | 22 | AA1981 |
| 3 | 2329 | 89.0 | 495 | 21 | AA1982 |
| 4 | 2329 | 89.0 | 495 | 22 | AA1983 |
| 5 | 2329 | 89.0 | 536 | 22 | AA1984 |
| 6 | 2329 | 89.0 | 536 | 22 | AA1985 |
| 7 | 2322 | 88.7 | 495 | 21 | AA1986 |
| 8 | 1702 | 65.0 | 495 | 21 | AA1987 |
| 9 | 1544 | 59.0 | 314 | 22 | AA1988 |
| 10 | 1232 | 47.1 | 277 | 22 | AA1989 |
| 11 | 1217 | 46.5 | 309 | 22 | AA1990 |

| | | | | | | |
|----|-------|------|-----|----|----------|--------------------|
| 12 | 991.5 | 37.9 | 559 | 22 | ABB58701 | Drosophila melanog |
| 13 | 985.5 | 37.7 | 502 | 22 | ABB60525 | Drosophila melanog |
| 14 | 985.5 | 37.7 | 502 | 22 | ABB65873 | Drosophila melanog |
| 15 | 956.5 | 36.5 | 582 | 22 | AAW79273 | Human protein SEQ |
| 16 | 946.5 | 36.2 | 516 | 22 | ABB67013 | Drosophila melanog |
| 17 | 941 | 36.0 | 529 | 22 | ABB63684 | Drosophila melanog |
| 18 | 932 | 35.6 | 567 | 22 | AAO13870 | Human polypeptide |
| 19 | 931 | 35.6 | 560 | 17 | AAW05148 | Human brain sodium |
| 20 | 931 | 35.6 | 560 | 19 | AAW70500 | Human sodium-lithi |
| 21 | 895 | 34.2 | 194 | 22 | AAW25685 | Human protein sequ |
| 22 | 859.5 | 32.8 | 481 | 22 | ABB64204 | Drosophila melanog |
| 23 | 845 | 32.3 | 479 | 22 | ABB62841 | Drosophila melanog |
| 24 | 835.5 | 31.9 | 496 | 22 | ABB61407 | Drosophila melanog |
| 25 | 819 | 31.3 | 576 | 20 | AAW88523 | Eat-4 protein amin |
| 26 | 798 | 30.5 | 465 | 22 | ABB64710 | Drosophila melanog |
| 27 | 787 | 30.1 | 512 | 22 | ABB60925 | Drosophila melanog |
| 28 | 784.5 | 30.0 | 497 | 22 | ABB70142 | Drosophila melanog |
| 29 | 767.5 | 29.3 | 462 | 22 | ABB70143 | Drosophila melanog |
| 30 | 760 | 29.0 | 493 | 22 | ABB59580 | Drosophila melanog |
| 31 | 744 | 28.4 | 512 | 21 | AAW29915 | Arabidopsis thalia |
| 32 | 744 | 28.4 | 519 | 21 | AAW29914 | Arabidopsis thalia |
| 33 | 737 | 28.2 | 475 | 22 | ABB67155 | Drosophila melanog |
| 34 | 737 | 28.2 | 491 | 22 | ABB70144 | Drosophila melanog |
| 35 | 729 | 27.9 | 395 | 21 | AAW29916 | Arabidopsis thalia |
| 36 | 715.5 | 27.3 | 560 | 22 | ABB59401 | Drosophila melanog |
| 37 | 697 | 26.6 | 449 | 22 | ABB58659 | Drosophila melanog |
| 38 | 696.5 | 26.6 | 524 | 22 | ABB64935 | Drosophila melanog |
| 39 | 672 | 25.7 | 436 | 19 | AAW78919 | Human haemochromat |
| 40 | 651 | 24.9 | 343 | 22 | ABB12999 | Novel human diagno |
| 41 | 618 | 23.6 | 495 | 22 | ABB64689 | Drosophila melanog |
| 42 | 611 | 23.3 | 124 | 22 | AAW81307 | Human AFP protein |
| 43 | 608 | 23.2 | 256 | 22 | ABG02032 | Novel human diagno |
| 44 | 603.5 | 23.1 | 401 | 19 | AAW78920 | Human haemochromat |
| 45 | 603.5 | 23.1 | 401 | 19 | AAW69971 | Human sodium-depen |

ALIGNMENTS

| | |
|--------|---|
| RESULT | 1 |
| AA1980 | AA1980 |
| ID | AA1980 standard; Protein; 495 AA. |
| XX | |
| AC | AA1980 |
| XX | |
| DT | 31-MAY-2000 (first entry) |
| XX | |
| DE | Sheep GBS toxin receptor (SP55). |
| XX | |
| KW | Sheep GBS toxin receptor; group B beta-haemolytic streptococci; SP55; |
| KW | pathological vascularisation; cancer metastases; angiogenesis; sheep; |
| KW | neovascularisation; reperfusion injury; scarring; keloid; |
| KW | chronic inflammation; rheumatoid arthritis; psoriasis; neural injury; |
| KW | endothelial cell proliferation; antibacterial; anticancer; |
| KW | anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic. |
| XX | |
| OS | Ovis sp. |
| XX | |
| FH | Key |
| FT | Domain |
| FT | Location/Qualifiers |
| FT | 226...252 |
| FT | /note= "Outer boundary of transmembrane domain" |
| FT | 232...248 |
| FT | /note= "Inner boundary of transmembrane domain" |
| FT | 365...389 |
| FT | /note= "Outer boundary of transmembrane domain" |
| FT | 369...385 |
| FT | /note= "Inner boundary of transmembrane domain" |
| FT | 456...479 |
| FT | /note= "Outer boundary of transmembrane domain" |
| FT | 458...474 |
| FT | /note= "Inner boundary of transmembrane domain" |
| FT | 135...157 |
| FT | /note= "Outer boundary of transmembrane domain" |

| | | | |
|----------|-------------------------------------|---|-----|
| Db | 61 | svalvdmvdsntaktndrtsyecahsapikvlhnqtgkkyrwdactggwllgsffgyi | 120 |
| Qy | 121 | ITQIPGGYVASRSGKLLGFGIFATAIFTLTPLAADFGVALVALRALEGEGVTYP | 180 |
| Db | 121 | itqipggyvasrsgklllgfifataiftltplaaadfgvalvalraleglgegvtyp | 180 |
| Qy | 181 | AMHAMSSWAPPLERSKLLSISYAGAQLTGVVSLPLSGVICYMMNWTYVYFFGIVGIIW | 240 |
| Db | 181 | amhamsswapplersskllsisyagaqlgtvvsplsgvicymmwnwtvyffgivgiw | 240 |
| Qy | 241 | FILWICLVSDTPTHKTIPTPYEKEYILLSKKNQSSQKSVWPIMLKSLPLWAIIVVAHFS | 300 |
| Db | 241 | filwiclvsdtpethktitpyekeyillskknqlssqksvpwipmlkslplwaiivvahfs | 300 |
| Qy | 301 | YNWTFYTLTLPTYMKEVLRNIOENGFLSAVPYLGWCMLCSGOAADNLRARNWFST | 360 |
| Db | 301 | ynwtfytltltpymkevrlrniengflsavpylgwcmllsgqaadnlararnwfst | 360 |
| Qy | 361 | LWVRRVFLSILGIMGPAIFLVAAGFIGCDYSLAVAFITISTTLGGFCSSGFSINHLDIAPS | 420 |
| Db | 361 | lwvrrvflsligmipailvaagfigcdyslavafitisttlggfcsgfsinhldiaps | 420 |
| Qy | 421 | YAGILLGINTFATIPGMIGPIIARSLTPENTIGEQTVFCIAAINVFGAIFFTLFAKG | 480 |
| Db | 421 | yagillgintfatipgmigpiiarsltpeintigeqtvfciaaainvfgaifftlfakg | 480 |
| Qy | 481 | EVQNWAISDHQGHNRN | 495 |
| Db | 481 | evqnwaishdqghnrn | 495 |
| RESULT 2 | | | |
| AAE06519 | | | |
| ID | AAE06519 standard; Protein; 495 AA. | | |
| XX | AC | AAE06519; | |
| XX | DT | 16-OCT-2001 (first entry) | |
| XX | DE | Sheep group B beta-haemolytic Streptococci toxin receptor (SP55) protein. | |
| XX | KW | Sheep; group B beta-haemolytic Streptococci toxin receptor; GBS; SP55; cytostatic; vulnary; antiatherosclerotic; osteopathic; vasotropic; | |
| XX | KW | prevention; attenuation; pathoangiogenic condition; cancer; scar; | |
| XX | KW | wound healing; gliosis; nerve injury; chronic wound; reperfusion injury; | |
| XX | KW | keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis; | |
| XX | OS | Ovis sp. | |
| XX | PH | Key | |
| XX | FT | Region | |
| XX | FT | /label= p55a_immunogenic_peptide | |
| XX | FT | /note= "Fragment of extracellular domain of GBS toxin receptor" | |
| XX | FT | Region | |
| XX | FT | /label= p56a_immunogenic_peptide | |
| XX | FT | /note= "Region of high hydrophilicity" | |
| XX | FT | Region | |
| XX | FT | /note= "Region of high hydrophilicity" | |
| XX | FT | Region | |
| XX | FT | /note= "Region of high hydrophilicity" | |
| XX | FT | Region | |
| XX | FT | /label= p57a_immunogenic_peptide | |
| XX | FT | /note= "Fragment of intracellular domain of GBS toxin receptor" | |
| XX | PN | WO200156598-A2. | |
| XX | XX | 09-AUG-2001. | |
| XX | PD | | |

| | | | | | |
|----|--|---|-------------------|-----------|-------------|
| PF | 02-FEB-2001; 2001WO-US03662. | | | | |
| XX | | | | | |
| PR | 02-FEB-2000; 2000US-0179870. | | | | |
| XX | | | | | |
| PA | (UYVA-) UNIV VANDERBILT. | | | | |
| XX | | | | | |
| PI | Hellerqvist CG; | | | | |
| XX | | | | | |
| DR | WPI; 2001-488844/53. | | | | |
| DR | N-PSDB; AAD10326. | | | | |
| XX | | | | | |
| PT | Preventing or attenuating pathoangiogenic conditions e.g. cancer, | | | | |
| PT | chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by | | | | |
| PT | administering group B beta-hemolytic Streptococci toxin receptor or its | | | | |
| PT | fragment - | | | | |
| XX | | | | | |
| PS | Claim 8; Page 50-52; 52pp; English. | | | | |
| XX | | | | | |
| CC | The present sequence is group B beta-haemolytic Streptococci (GBS) toxin receptor protein, SP55 from sheep. The present invention relates to a method for preventing or attenuating a patho-angiogenic condition in a mammal which comprises administering to the mammal one or more GBS toxin receptors or their immunogenic fragments to induce or maintain an immune response to one of GBS toxin receptors. The method is useful for preventing or ameliorating pathoangiogenic conditions such as cancer, scarring during wound healing, gliosis during repair of nerve injury, chronic wounds, keloids, reperfusion injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and psoriasis in mammals. The proteins of the invention are also used as vaccines. | | | | |
| XX | | | | | |
| SQ | Sequence | 495 AA; | | | |
| | Query Match | 100.0%; | Score 2617; | DB 22; | Length 495; |
| | Best Local Similarity | 100.0%; | Pred. No. 5e-276; | | |
| | Matches 495; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| QY | 1 | MKSPVSDLAPSDGEGSDRTPLLRAPRAEPAPVCCSARYNLAFISFFGFVLYSLRVNL | 60 | | |
| DB | 1 | mkspsvdlapsdgegsdrtpllqrapraepapvccsarynlafisfgffvlyslrvnl | 60 | | |
| QY | 61 | SVALVDMVDSNTAKDNRTSYECAHSAPIKVLHNQTGKKYRWDATQGWILGSFFGYI | 120 | | |
| DB | 61 | svalvdmvdsntaktndrtsyecahsapikvlhnqtgkkyrwdactggwllgsffgyi | 120 | | |
| QY | 121 | ITQIPGGYVASRSGKLLGFGIFATAIFTLTPLAADFGVALVALRALEGEGVTYP | 180 | | |
| DB | 121 | itqipggyvasrsgklllgfifataiftltplaaadfgvalvalraleglgegvtyp | 180 | | |
| QY | 181 | AMHAMSSWAPPLERSKLLSISYAGAQLTGVVSLPLSGVICYMMNWTYVYFFGIVGIIW | 240 | | |
| DB | 181 | amhamsswapplerskllsisyagaqlgtvvsplsgvicymmwnwtvyffgivgiw | 240 | | |
| QY | 241 | FILWICLVSDTPTHKTIPTPYEKEYILLSKKNQSSQKSVWPIMLKSLPLWAIIVVAHFS | 300 | | |
| DB | 241 | filwiclvsdtpethktitpyekeyillskknqlssqksvpwipmlkslplwaiivvahfs | 300 | | |
| QY | 301 | YNWTFYTLTLPTYMKEVLRNIOENGFLSAVPYLGWCMLCSGOAADNLRARNWFST | 360 | | |
| DB | 301 | ynwtfytltltpymkevrlrniengflsavpylgwcmllsgqaadnlararnwfst | 360 | | |
| QY | 361 | LWVRRVFLSILGIMGPAIFLVAAGFIGCDYSLAVAFITISTTLGGFCSSGFSINHLDIAPS | 420 | | |
| DB | 361 | lwvrrvflslgimgpailvaagfigcdyslavafitisttlggfcsgfsinhldiaps | 420 | | |
| QY | 421 | YAGILLGINTFATIPGMIGPIIARSLTPENTIGEQTVFCIAAINVFGAIFFTLFAKG | 480 | | |
| DB | 421 | yagillgintfatipgmigpiiarsltpeintigeqtvfciaaainvfgaifftlfakg | 480 | | |
| QY | 481 | EVQNWAISDHQGHNRN | 495 | | |
| DB | 481 | evqnwaishdqghnrn | 495 | | |

RESULT 3
 AAY45087
 ID AAY45087 standard; Protein; 495 AA.
 AC AAY45087;
 XX 31-MAY-2000 (first entry)
 DT Partial human GBS toxin receptor (HP55).
 DE Human GBS toxin receptor; group B beta-haemolytic streptococci; HP55;
 XX pathological vascularisation; cancer metastases; angiogenesis;
 KW neovascularisation; reperfusion injury; scarring; keloid;
 KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
 KW endothelial cell proliferation; antibacterial; anticancer;
 KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Region 14..20
 FT /note= "Region of high hydrophilicity used to design
 FT antigenic peptide"
 FT Region 76..80
 FT /note= "Region of high hydrophilicity used to design
 FT antigenic peptide"
 FT Region 26..30
 FT /note= "Region of high hydrophilicity used to design
 FT antigenic peptide"
 XX WO200005375-A1.
 XX 03-FEB-2000.
 PD 22-JUL-1999; 99WO-US16676.
 PF 22-JUL-1998; 98US-0093843.
 XX (UYVA-) UNIV VANDERBILT.
 XX Hellerqvist CG, Fu C;
 DR WPI; 2000-205377/18.
 DR N-PSDB; AA250875.
 PT New polynucleotide encoding mammalian receptor for streptococcus toxin,
 PT useful for diagnosis and treatment of, e.g. pneumonia in neonates -
 XX Claim 9; Page 22; 109pp; English.
 CC The present sequence is partial human GBS (group B beta-haemolytic
 CC streptococci) toxin receptor (HP55). GBS toxin receptor is an integral
 CC protein with seven transmembrane domains. Expression vectors comprising
 CC the coding region can be transformed into host cells to express GBS
 CC toxin receptor and its fragments. Detecting the receptor in tissues is
 CC used to diagnose pathological vascularisation, e.g. for detecting cancer
 CC metastases. GBS toxin receptors are useful for treating conditions
 CC associated with pathological angiogenesis or neovascularisation
 CC (specifically cancer, reperfusion injury, scarring during wound healing,
 CC keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
 CC neural injury), and to raise specific antibodies used for treating early
 CC onset disease. Inhibitors of this receptor are useful for treating
 CC pathological or hypoxia-induced endothelial cell proliferation and
 CC migration.
 XX
 SQ Sequence 495 AA;
 Query Match 89.0%; Score 2329; DB 21; Length 495;
 Best Local Similarity 86.3%; Pred. No. 1.2e-244;
 Matches 427; Conservative 39; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKSPVSDLAPSDGEGSDRTPLQRAPRAEPAPVCCSARYNLAFSLFFGFFVLYSLRVNL 60
 DB 1 mcsprvdlarndgeestdrtpllpqapraeapvccsarynlailaffgfviyalrvnl 60
 QY 61 SVALVDMVDSNTAKDNRTSYECAHSAPIKVLHNQTKKRYWDATQGWILGSPFYGI 120
 DB 61 svalvdmvdsntlednrtskacpehsapikvhhnqtgkkyqwdactggwiigsfgyi 120
 QY 121 ITQIPGGYVASRSGKLLGFGIFATAIFTPLAADFVGVALVALRALEGLEGVTVTP 180
 DB 121 itqipggyvasrsgklllgfifataiftplaaadfgvalvalraleglegvtfp 180
 QY 181 AMHAMWSSWAPPLERSKLLSISYAGAQLTGVVSLPSGVICYMNMVYVYFFGVIGIIV 240
 DB 181 amhamwsswapplerskllsisyagaqlgtvslpslgicymnmvtyvyffgtigifw 240
 QY 241 FILWICLVSDTETHKTIPTYKEYILSLKQLSSQKSVPMKLSPLWAIIVVAHFS 300
 DB 241 filwiclvsdtpqkhhkrihsykeyilslrnqlssqksvpwvpilksiplwaivvahfs 300
 QY 301 YNWTFTLTLLPTYMKVLENIQENGLSAVPYLGCWLCMLTSGQAADNLRARNFST 360
 DB 301 ynwtfytltltpymkellrinvqengflsslpylgswicmlsgqaadnlrakwnfst 360
 QY 361 LMWRRVFLSIGMIGPAIFLVAAGFIGCDYSLAVAFLTISTTLGGFCSSGFSINHLDIAPS 420
 DB 361 lcwrrifsligmigpavflvaagfigcdyslavafiltsttlggfcsgfsinhldiaps 420
 QY 421 YAGILLGITNTFATPGMIGPIIARSLTPENTIGEQTVFCIAAINVFGAIFTLFAKG 480
 DB 421 yagillgitntfatpgmvgpviaksltpdntvgewqtvfyiaaainvfgaifftlfakg 480
 QY 481 EVQNWAISSDHOGRN 495
 DB 481 evqnwaindhghrh 495
 RESULT 4
 AAB66967
 ID AAB66967 standard; protein; 495 AA.
 XX AC AAB66967;
 XX 18-APR-2001 (first entry)
 DT Human AST.
 DE Human AST.
 XX Human; AST; nootropic; immunotropic; gene therapy; Salla disease;
 KW anion and sugar transporter; anion-cation symporter;
 KW sialic acid transporter.
 XX Homo sapiens.
 OS EP1069184-A1.
 PN 17-JAN-2001.
 PD 16-JUL-1999; 99EP-0202341.
 XX 16-JUL-1999; 99EP-0202341.
 PA (ALKU) AKZO NOBEL NV.
 DR WPI; 2001-193090/20.
 DR N-PSDB; AAF55900.
 XX New human transporter gene implicated in Salla disease and lysosomal
 PT sialic acid transport, useful in assays for identifying new drugs, or
 PT diagnosing sialic acid transport defects related to mutations in the
 XX transporter gene -
 PS Claim 1; Page 10-12; 20pp; English.
 XX

| | | | | | | | | | | |
|--|--|---|-----|--|--|--|--|--|--|--|
| CC | The present sequence is human Anion and Sugar Transporter (AST) protein sequence. AST has significant homology with several members of the anion-cation symporter (ACS) family of transporters. AST is implicated in Salla disease, and is useful in screening assays for identifying new drugs. Compounds identified via AST screening is useful for preparing a pharmaceutical suitable as an activator or inhibitor of a sialic acid transporter protein. The pharmaceutical may be used in sialic acid associated diseases and CNS/immune related disorders. | | | | | | | | | |
| XX | | | | | | | | | | |
| SQ | Sequence 495 AA; | | | | | | | | | |
| Query Match 89.0%; Score 2329; DB 22; Length 495; | | | | | | | | | | |
| Best Local Similarity 86.3%; Pred. No. 1.2e-244; | | | | | | | | | | |
| Matches 427; Conservative 39; Mismatches 29; Indels 0; Gaps 0; | | | | | | | | | | |
| QY | 1 | MKSPVSLAPSDGEESDRTPLLQRAPEAPVCCSARYNLAFSGFEFVLYLRVNL | 60 | | | | | | | |
| Db | 1 | mrspvrdlarndgeesdrplpgpraeeapvccsarynlalaffgfvlvaylrnl | 60 | | | | | | | |
| QY | 61 | SVALVDVDSNTTAKDNRTSYECAHSAPIKVLHNTGKKYRMDAETQGWILGSFFYGYI | 120 | | | | | | | |
| Db | 61 | svalvdmvdsnttlednrtskacpehsapikvhhnqtgkkyqdaetqgwilgsffgyi | 120 | | | | | | | |
| QY | 121 | ITQIPGGYVASRGGKLLGFGIFATAFTLFTPLAADFGVGALVALRLEGIGVYTP | 180 | | | | | | | |
| Db | 121 | itqipggyvasrkigkmlilgflgtavltlftpiaadlvgvplvlrallengivgtfp | 180 | | | | | | | |
| QY | 181 | AMHAMSSWAPPLERSKLLSISVAGAQGVVSLPLSGVTCYVMNTYVYFFFGIVGIW | 240 | | | | | | | |
| Db | 181 | amhamsswapplerskllsisvagaqgvcvslplsgvtycmntvtyvfffgtvgifw | 240 | | | | | | | |
| QY | 241 | FILWICLVSDTPETHRTITPYKEYILSSLKNQLSSQKSPWPIPLMLKSLPLWAIIVVAHFS | 300 | | | | | | | |
| Db | 241 | flilwlvsdtpqkhrishyekeyilsslirnlqslsqkspwpiplsplwaiivvahfs | 300 | | | | | | | |
| QY | 301 | YNWTFYTLTLPTKYKEVLRFNIQENGLSVPYLGWCMLTSLGQADNLRARNFST | 360 | | | | | | | |
| Db | 301 | ynwtfytltlptymkelirfnvqngfvlssipyilgswlcmilsggaadnkrakwnfst | 360 | | | | | | | |
| QY | 361 | LWRRVFSLIGMTGPAIFLVAAGFICGDSYLAFAFLTISTLGGFCSSGFSINHLDIAPS | 420 | | | | | | | |
| Db | 361 | lvrrifsligmipgavflvaagfigcdyslavafitistlggfcssgfsinhldiaps | 420 | | | | | | | |
| QY | 421 | YAGILGITNTFATIPGMIPTIARSILTPENTIGEQTVFCIAAANVFCIAIFFTLFAKG | 480 | | | | | | | |
| Db | 421 | yagilgintntfatipgmvgpvlaaksltpdntvgewqtvfyaaaainvfgaifftlfakg | 480 | | | | | | | |
| QY | 481 | EVONWAIISDHQHRN | 495 | | | | | | | |
| Db | 481 | evqpnwalndhbghrh | 495 | | | | | | | |
| RESULT 5 | | | | | | | | | | |
| AA45089 | | | | | | | | | | |
| ID | AA45089 standard; Protein; 536 AA. | | | | | | | | | |
| XX | | | | | | | | | | |
| AC | AA45089; | | | | | | | | | |
| XX | | | | | | | | | | |
| DT | 31-MAY-2000 (first entry) | | | | | | | | | |
| XX | | | | | | | | | | |
| DE | Human GBS toxin receptor (HP59). | | | | | | | | | |
| XX | | | | | | | | | | |
| KW | Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59; | | | | | | | | | |
| KW | pathological vascularisation; cancer metastases; angiogenesis; | | | | | | | | | |
| KW | neovascularisation; reperfusion injury; scarring; keloid; | | | | | | | | | |
| KW | chronic inflammation; rheumatoid arthritis; psoriasis; neural injury; | | | | | | | | | |
| KW | endothelial cell proliferation; antibacterial; anticancer; | | | | | | | | | |
| XX | anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic. | | | | | | | | | |
| OS | Homo sapiens. | | | | | | | | | |
| XX | | | | | | | | | | |
| FH | Key | Location/Qualifiers | | | | | | | | |

| | | | |
|----|---------------|----------|--|
| FT | Modified-site | 23..26 | /note= "Putative amidation site" |
| FT | Modified-site | 138..141 | /note= "Putative amidation site" |
| FT | Modified-site | 100..103 | /note= "Asn is N-glycosylated" |
| FT | Modified-site | 112..115 | /note= "Asn is N-glycosylated" |
| FT | Modified-site | 118..121 | /note= "Asn is N-glycosylated" |
| FT | Modified-site | 136..139 | /note= "Asn is N-glycosylated" |
| FT | Modified-site | 286..269 | /note= "Asn is N-glycosylated" |
| FT | Modified-site | 343..346 | /note= "Asn is N-glycosylated" |
| FT | Modified-site | 398..401 | /note= "Asn is N-glycosylated" |
| FT | Modified-site | 297..300 | /label= CAMP_phospho_site |
| FT | Modified-site | 113..116 | /note= "Putative phosphorylation site" |
| FT | Modified-site | 114..117 | /label= CK2_phospho_site |
| FT | Modified-site | 300..303 | /note= "Putative phosphorylation site" |
| FT | Modified-site | 493..496 | /label= CK2_phospho_site |
| FT | Modified-site | 66..71 | /note= "Putative phosphorylation site" |
| FT | Modified-site | 167..172 | /note= "Putative myristylation site" |
| FT | Modified-site | 183..188 | /note= "Putative myristylation site" |
| FT | Modified-site | 213..218 | /note= "Putative myristylation site" |
| FT | Modified-site | 246..251 | /note= "Putative myristylation site" |
| FT | Modified-site | 250..255 | /note= "Putative myristylation site" |
| FT | Modified-site | 378..383 | /note= "Putative myristylation site" |
| FT | Modified-site | 427..432 | /note= "Putative myristylation site" |
| FT | Modified-site | 444..449 | /note= "Putative myristylation site" |
| FT | Modified-site | 464..469 | /note= "Putative myristylation site" |
| FT | Modified-site | 468..473 | /note= "Putative myristylation site" |
| FT | Modified-site | 23..25 | /label= PKC_phospho_site |
| FT | Modified-site | 58..60 | /note= "Putative phosphorylation site" |
| FT | Modified-site | 78..80 | /label= PKC_phospho_site |
| FT | Modified-site | 120..122 | /note= "Putative phosphorylation site" |
| FT | Modified-site | 138..140 | /label= PKC_phospho_site |
| FT | Modified-site | 310..312 | /note= "Putative phosphorylation site" |
| FT | Modified-site | 310..312 | /label= PKC_phospho_site |
| FT | Modified-site | 310..312 | /note= "Putative phosphorylation site" |

Modified-site 317..320
/label= PKC_phospho_site
/note= "Putative phosphorylation site"
WO200005375-A1.
03-FEB-2000.
22-JUL-1999; 99WO-US16676.
22-JUL-1998; 98US-0093843.
(UYVA-) UNIV VANDERBILT.
Hellerqvist CG, Fu C;
WPI; 2000-205377/18.
N-PSDB; AA250879.
New polynucleotide encoding mammalian receptor for streptococcus toxin,
useful for diagnosis and treatment of, e.g. pneumonia in neonates -
Claim 10; Page 93-95; 109pp; English.
The present sequence is partial human GBS (group B beta-haemolytic
streptococci) toxin receptor (HP55). GBS toxin receptor is an integral
protein with seven transmembrane domains. Expression vectors comprising
the coding region can be transformed into host cells to express GBS
toxin receptor and its fragments. Detecting the receptor in tissues is
used to diagnose pathological vascularisation, e.g. for detecting cancer
metastases. GBS toxin receptors are useful for treating conditions
associated with pathological angiogenesis or neovascularisation
(specifically cancer, reperfusion injury, scarring during wound healing,
keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
neural injury), and to raise specific antibodies used for treating early
onset disease. Inhibitors of this receptor are useful for treating
CC pathological or hypoxia-induced endothelial cell proliferation and
CC migration.
Sequence 536 AA;
Query Match 89.0%; Score 2329; DB 21; Length 536;
Best Local Similarity 86.3%; Pred. No. 1.4e-244;
Matches 427; Conservative 39; Mismatches 29; Indels 0; Gaps 0;
QY 1 MKSPVSLAPSDGEGSDRPTLQRAPAPVCCSARYNLAPLSPFGFVLYSLRVNL 60
Db mrsprdlarndgeestdrtpllpgpraeapvccsarynlailatgffivyalrvnl 101
QY 61 SVALVDMVDNTAKDNRTSYECAEHSAPIKVLHNOTGKKYRWDATQGWILGSEFFGYI 120
Db 102 svalvdmvdnttdrntskacpelsapikvvhmctgkkygwdatcgwllgsffgyi 161
QY 121 ITQIPGYVASRGGKLLGLGIFATAIFTPLAADFGVAGALVALRALEGEGYTYP 180
Db 162 itqipgyvasrkgkmlgflgltavltlftpiaadlcvglvrlraleleggytft 221
QY 181 AMHAMWSSWAPPLERSKLLSISAGAGLGVTVSLPSGVICYNNWTVYFFPGVIGIIV 240
Db 222 amhamwsswapplerskllsisagaglgvtvslpsgvicynnwmtvyffgtgfgw 281
QY 241 FILWICLVSDPTFKHTITPYEKEYILSLKLNLSQKSPWIPMLKSLPLWAIIVAHFS 300
Db 282 flwlvdsdtpqkhrisyekeyilslrnlqslsqkspwipmlksplwaiivahfs 341
QY 301 YNNTFTYLLPTMYKEVLFNFIQENGFLSVPYLGWCMLISGQAADNLARNWFST 360
Db 342 ynwtyftlltptymkeilfnvqengflsslpylgswclmllsgqaadnlrknwfst 401
QY 361 LWRRVPSLLGMPAIFLVAAGFICGDSYLAVALFTISTTLGFCSSGFSINHLDIAPS 420
Db 402 lcwrrfslgmpavflvaagfigcdyslavafitisttlgfcsgfsinhldiaps 461

QY 421 YAGILLGITNTFATIPGMIGPIIARSITPENTIGEMOTVFCIAAIAINVGAIFFTLFAK 480
Db 462 yagillgitntfatipgmvgpviaksltpdntvgewqtvfyaaainvfgaifftlfak 521
QY 481 EVQNWAISSDHQHRN 495
Db 522 evqnwaindhghrh 536
RESULT 6
AAE06518
ID AAE06518 standard; Protein; 536 AA.
XX AC AAE06518;
XX DT 16-OCT-2001 (first entry)
XX Human group B beta-haemolytic Streptococci toxin receptor (HP59) protein.
DE Human; group B beta-haemolytic Streptococci toxin receptor; GBS; HP59;
XX cytostatic; vulnary; antiatherosclerotic; osteopathic; vasotropic;
KW prevention; attenuation; pathoangiogenic condition; cancer; scar;
KW wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;
KW keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;
KW vaccine.
XX Homo sapiens.
XX Key Location/Qualifiers
FH Region 8..28 /label= Hab3_immunogenic_peptide
FT Region 49..63 /label= Hab1_immunogenic_peptide
FT Region 49..76 /label= Hab4_immunogenic_peptide
FT Region 112..125 /label= Hab2_immunogenic_peptide
XX WO200156598-A2.
XX 09-AUG-2001.
XX 02-FEB-2001; 2001WO-US03662.
XX 02-FEB-2000; 2000US-0179870.
XX (UYVA-) UNIV VANDERBILT.
XX Hellerqvist CG;
XX WPI; 2001-488844/53.
XX N-PSDB; AAD10325.
XX Preventing or attenuating pathoangiogenic conditions e.g. cancer,
PT chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by
PT administering group B beta-hemolytic Streptococci toxin receptor or its
PT fragment -
XX Claim 4; Page 45-47; 52pp; English.
XX The present sequence is full length group B beta-haemolytic Streptococci
CC (GBS) toxin receptor protein, HP59 from human. The present invention
CC relates to a method for preventing or attenuating a patho-angiogenic
CC condition in a mammal which comprises administering to the mammal one
CC or more GBS toxin receptors or their immunogenic fragments to induce
CC or maintain an immune response to one of GBS toxin receptors. The
CC method is useful for preventing or ameliorating pathoangiogenic
CC conditions such as cancer, scarring during wound healing, gliosis
CC during repair of nerve injury, chronic wounds, keloids, reperfusion
CC injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and
CC psoriasis in mammals. The proteins of the invention are also used
CC as vaccines.

[illegible]

```

|||||
421 yagillgitntfatipgmvgpviaakstlpdntvgewqtvfyiaaaainvgaiftlifakg 480
QY 481 EVQVNAISDHQHRN 495
DB 481 evqvnalndhhgrh 495

RESULT 8
AAY45090
ID AAY45090 standard; Protein; 495 AA.
AC AAY45090;
DT 31-MAY-2000 (first entry)
DE Human/Sheep consensus GBS toxin receptor.
KW Human/Sheep consensus GBS toxin receptor; GBS;
KW group B beta-haemolytic streptococci; pathological vascularisation;
KW cancer metastases; angiogenesis; neovascularisation; reperfusion injury;
KW scarring; keloid; chronic inflammation; rheumatoid arthritis; psoriasis;
KW neural injury; endothelial cell proliferation; antibacterial;
KW anticancer; anti-angiogenic; anti-inflammatory; anti-arthritis;
KW anti-psoriatic.
OS Homo sapiens.
OS Ovis sp.
FH Key Location/Qualifiers
FT Protein 1..495
FT /note= "All the X's in the sequence correspond to
FT non-consensus sites"
XX
PN WO200005375-A1.
XX
XX 03-FEB-2000.
XX
XX 22-JUL-1999; 99WO-US16676.
XX
XX 22-JUL-1998; 98US-0093843.
XX
XX (UYVA-) UNIV VANDERBILT.
XX
XX Hellerqvist CG, Fu C;
XX
XX WPI; 2000-205377/18.
XX
XX N-PSDB; AA250880, AA250881.
XX
XX New polynucleotide encoding mammalian receptor for streptococcus toxin,
XX useful for diagnosis and treatment of, e.g. pneumonia in neonates -
XX
XX Disclosure; Page 98-100; 109pp; English.
XX
XX The present sequence is the human/sheep consensus GBS (group B beta
XX -haemolytic streptococci) toxin receptor. Expression vectors comprising
XX the coding region can be transformed into host cells to express GBS
XX toxin receptor and its fragments. Detecting the receptor in tissues is
XX used to diagnose pathological vascularisation, e.g. for detecting cancer
XX metastases. GBS toxin receptors are useful for treating conditions
XX associated with pathological angiogenesis or neovascularisation
XX (specifically cancer, reperfusion injury, scarring during wound healing,
XX keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
XX neural injury), and to raise specific antibodies used for treating early
XX onset disease. Inhibitors of this receptor are useful for treating
XX pathological or hypoxia-induced endothelial cell proliferation and
XX migration.
XX
XX Sequence 495 AA;

```

Query Match 65.0%; Score 1702; DB 21; Length 495;
 Best Local Similarity 67.8%; Pred. No. 2.7e-176;

```

Matches 335; Conservative 0; Mismatches 159; Indels 0; Gaps 0;
QY 1 MKSPVSDLAPSDGEGSDRTPLLQRAPEAPAPVCCSARYNLAFLSFFGFFVLSRVNL 60
DB 1 mxpxvdxaxxxgeexdrxxxxxaxrxxapccsarynaxlxfgffxxylxvnl 60
QY 61 SVALVDMVDSNTAKDNRTSYECAHSAPIKVLHNTGKGRWDAETQCWILGSPFYGYI 120
DB 61 vxxvxxmxdsxttxdnrxsxxcxehsapikvxxxqtkgkxwdaetqgwilxxfygyi 120
QY 121 ITQIPGGYVASRGGKLLGFGIFATAIFTLFTPLAADFVGVGALVALRALEGLGEGVTP 180
DB 121 itxipggyvasxxgxxlqgixxxaxxtlftpxaadxxgxxxxlxalexlgegtxp 180
QY 181 AMHAMSSWAPPLERSKLLISYAGAQLTGVVSLPLSGVICYMMNWTYVYFPGIIGIIV 240
DB 181 amhamsswappplersxlxxiyagaxlgtvslplsgvicymnwtvyvfxixxxgxxw 240
QY 241 FILWICLVSDTPETHKTITPYEKEYTLSSLKQLSOKSPVPIMLKSLPLWAIIVAHFS 300
DB 241 fxxwixlvxtpxxhxxkxxxxyekxllslxqxsqskvpxxxxxxlpwaxxvaxfs 300
QY 301 YNWTFTLLTLLPTYMKVLFNFIQENGLSAVPYLGWCMLSGQAADNLRARWNFT 360
DB 301 ynwftfytlllptxmkxxlrfnxqngflsxxpylxxwlcmlxgqaadnlarxwnfst 360
QY 361 LWVRVFLSLIGMIGPAIFLVAAGFTGCDYSLAVAFITITTLGGFCSSGFSINHLDIAP 420
DB 361 xxvxrxfslixmipgxxflxxxxxxgcyxlxxvflxistxlggfcsgfsinhxiaps 420
QY 421 YAGILGINTTATIPGMIGPIIARSLTPENTIGEMQTVFCIAAANVFCAIFFTLFAG 480
DB 421 yagilgicnxfatixgmxxpxixsxtbntxgwxqxxfxaaaxnxfagixftlifakg 480
QY 481 EVQVNAISDHQHR 494
DB 481 exqvnxxxxdhxghr 494

RESULT 9
AAG65238
ID AAG65238 standard; Protein; 314 AA.
XX
XX AC AAG65238;
XX
XX DT 20-NOV-2001 (first entry)
XX
XX DE Human sodium dependent phosphate co-transfer protein 35.
XX
XX KW Human; sodium dependent phosphate co-transfer protein 35;
XX KW hypophosphaturia; hypercalcaemia; hypophosphataemic rickets;
XX KW nephritis; gene therapy.
XX
XX OS Homo sapiens.
XX
XX PN CN1298882-A.
XX
XX PD 13-JUN-2001.
XX
XX PF 06-DEC-1999; 99CN-0124217.
XX
XX PR 06-DEC-1999; 99CN-0124217.
XX
XX (SHEN-) SHENGYUAN GENE DEV CO LTD SHANGHAI.
XX
XX PI Mao Y, Xie Y;
XX
XX WPI; 2001-503367/56.
XX
XX N-PSDB; AAH79234.
XX
XX Human Na-dependent phosphate cotransporter 35 and its coding sequence -
XX Claim 1; Page 21(Disclosure); 28pp; Chinese.
PS

```

XX The present invention provides the protein and coding sequences of human
CC sodium dependent phosphate co-transfer protein 35. The sequences can be
CC used in the treatment of hypophosphaturia, hypercalcaemia,
CC hypophosphataemic rickets and nephritis. The present sequence is the
CC protein of the invention.
XX
SQ Sequence 314 AA;

Query Match 59.0%; Score 1544; DB 22; Length 314;
Best Local Similarity 89.2%; Pred. No. 2.3e-159;
Matches 280; Conservative 24; Mismatches 10; Indels 0; Gaps 0;

Qy 182 MHAMSSWAPPLERSKLLSISYAGAQLTGVVSLPLSGVICYMMNTVYVFFGIVGIWF 241

Db 1 mhamswwapplerskllsisyagaqlgtvislpslglicymntvtyvffgtigifw 60

Qy 242 ILWICLVSDTETHKTIPTPYEKEYILLSSKLNQSSQKSPWIPMLKSLPLWAIIVVAHFSY 301

Db 61 llwlvwsvdtpqkhrishyekeyillsrnqlssqkspwvplksiplwlvvvhfsy 120

Qy 302 NWTFTLLTLPTMYKEVLRNIOENGLSAVPYLGWCLMTLSGQAADNLRARNFSTL 361

Db 121 nwtftlltlptymkeilrfnvqngfssipygswlcmllsqgaadnhrakwnfstl 180

Qy 362 WYRRVFSLIGMIPAIFLVAAGFIGCDYSLAVAFLTISTTLGGFCSSGFSINHLDIAPSY 421

Db 181 cvrrifsligmipavflvaagfigcdyslavafitisttlggfcssgfsinhldiapsy 240

Qy 422 AGILLGINTFATPGMIGPIIARSLTPENIGEWQVFCIAAINVFGAIFFTLFAKGE 481

Db 241 agillgintfatpgmvgpviaksltpdntvgewqvfyaainvfgaifftlfakge 300

Qy 482 VQWVAISDHQGRN 495

Db 301 vqnwaindhghnrh 314

RESULT 10

AAM93914
ID AAM93914 standard; Protein; 277 AA.

XX
AC AAM93914;

XX
DT 06-NOV-2001 (first entry)

XX
DE Human polypeptide, SEQ ID NO: 4069.

XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.

XX
OS Homo sapiens.

XX
PN EP1130094-A2.

XX
PD 05-SEP-2001.

XX
PF 07-JUL-2000; 2000EP-0114089.

XX
PR 08-JUL-1999; 99JP-0194486.

PR 11-JAN-2000; 2000JP-0118774.

PR 02-MAY-2000; 2000JP-0183765.

XX
PA (HELI-) HELIX RES INST.

XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

XX
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX
XX WPI: 2001-524255/58.

XX
DR N-PSDB; AAK94876.

XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -

XX

PS Claim 8; SEQ ID NO 4069; 1380pp + sequence listing; English.

CC The invention relates to primers for synthesising full length cDNA

CC clones. 830 cDNA molecules encoding a human protein have been

CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA

CC molecules have been determined. Primers for synthesising the full length

CC cDNA are useful for clarifying the function of the protein encoded by

CC the cDNA. The full length clones were obtained by construction of full

CC length enriched cDNA libraries that were synthesised by the oligo-capping

CC method. The primers enable the production of the full length cDNA easily

CC without any special methods. The present sequence is a polypeptide

CC encoded by a full length human cDNA of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in CD-ROM format directly from EPO.

XX
SQ Sequence 277 AA;

Query Match 47.1%; Score 1232; DB 22; Length 277;

Best Local Similarity 81.9%; Pred. No. 1.9e-125;

Matches 227; Conservative 22; Mismatches 28; Indels 0; Gaps 0;

Qy 1 MKSPVSDIAPSDGEGSDRTPLLPQAPRAEPAPVCCSARYNLAFSLFFGVLSLVNL 60

Db 1 mrsprvdlarndgeestdrtpllpgapraeapvccsarynlalaffgfivayalrnl 60

Qy 61 SVALVDMVDSNTAKDNRTSYECASHSAPIKVLHNOTGKYRWDATOGWILGSFFYGYI 120

Db 61 svalvdmvdsntlednrtskacpshsapikvhhnqtgkqwdatogwvlgsgffyyi 120

Qy 121 ITQIPGGYVASKRSGKLLGFGIFATAFTLTPLAADPGVAGLVALRALEGLSGEVTP 180

Db 121 itqipggyvaskrgkllgfgilgtavltlftplaaadlgvpllvralleglsgvufp 180

Qy 181 AMHAMSSWAPPLERSKLLSISYAGAQLTGVVSLPLSGVICYMMNTVYVFFGIVGIW 240

Db 181 amhamsswappplerskllsisyagaqlgtvislpslglicymntvtyvffgtigifw 240

Qy 241 FILWICLVSDTETHKTIPTPYEKEYILLSLKNQLSSQ 277

Db 241 filwlvsvdtpqkhrishyekeyillslnqagvq 277

RESULT 11

AAM40745.

ID AAM40745 standard; Protein; 309 AA.

XX
AC AAM40745;

XX
DT 22-OCT-2001 (first entry)

XX
DE Human polypeptide SEQ ID NO 5676.

XX
KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;

XX
KW peripheral nervous system; neuropathy; central nervous system; CNS;

XX
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

XX
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

XX
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

XX
XX leukaemia.

OS
OS Homo sapiens.

XX
PN WO200153312-A1.

XX
PD 26-JUL-2001.

XX
PF 26-DEC-2000; 2000WO-US34263.

XX
PR 21-JAN-2000; 2000US-0488725.

XX
PR 25-APR-2000; 2000US-052317.

XX
PR 09-JUL-2000; 2000US-0598042.

XX
PR 19-JUL-2000; 2000US-0620312.

QY 441 PIIARSITPENTIGEWQTVFCIAAAINVECAIFFFTLFAKGEVONWA 486
||| :|| : ||| || ||| :| ||| :|| || ||
Db 461 piivgamtknksreewqyvfliaalvhyggvifayafasgekpwa 506

Search completed: July 15, 2002, 08:24:02
Job time: 107 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 08:22:58 ; Search time 32.66 Seconds
(without alignments)
1456.345 Million cell updates/sec

Title: US-09-776-865-4
Perfect score: 2617
Sequence: 1 MKSPVSLAPSDGEGSDRT.....LFAKGEVQNWAISHQHRN 495

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 968.5 | 37.0 | 493 | 2 G88553 | protein C38C10.2 [|
| 2 | 950 | 36.3 | 472 | 2 S28286 | hypothetical prote |
| 3 | 927 | 35.4 | 560 | 2 I59302 | brain specific Na+ |
| 4 | 819 | 31.3 | 563 | 2 T34650 | probable sodium-de |
| 5 | 819 | 31.3 | 576 | 2 H88548 | protein ZK512.6 [i |
| 6 | 809.5 | 30.9 | 465 | 2 A56410 | sodium/phosphate t |
| 7 | 802.5 | 30.7 | 465 | 2 S69915 | sodium-phosphate t |
| 8 | 792.5 | 30.3 | 573 | 2 T23589 | hypothetical prote |
| 9 | 791.5 | 30.2 | 467 | 2 A48916 | sodium phosphate t |
| 10 | 775.5 | 29.6 | 465 | 2 I39473 | Na+-dependent phos |
| 11 | 766.5 | 29.3 | 544 | 2 T24633 | hypothetical prote |
| 12 | 744 | 28.4 | 512 | 2 H84698 | hypothetical prote |
| 13 | 585 | 22.4 | 501 | 2 B9135 | protein F25G6.7 [i |
| 14 | 561.5 | 21.5 | 413 | 2 T01534 | hypothetical prote |
| 15 | 550.5 | 21.0 | 537 | 2 T45634 | hypothetical prote |
| 16 | 541.5 | 20.7 | 466 | 2 S40767 | hypothetical prote |
| 17 | 499 | 19.1 | 561 | 2 D84800 | hypothetical prote |
| 18 | 490.5 | 18.7 | 592 | 2 T25419 | hypothetical prote |
| 19 | 489 | 18.7 | 568 | 2 S44742 | C02C2.4 protein - |
| 20 | 471 | 18.0 | 530 | 2 T29418 | hypothetical prote |
| 21 | 450 | 17.2 | 380 | 2 T29248 | hypothetical prote |
| 22 | 444.5 | 17.0 | 516 | 2 T24729 | hypothetical prote |
| 23 | 428 | 16.4 | 499 | 2 T15201 | hypothetical prote |
| 24 | 420 | 16.0 | 506 | 2 T29968 | hypothetical prote |
| 25 | 419.5 | 16.0 | 543 | 2 T32496 | hypothetical prote |
| 26 | 395.5 | 15.1 | 493 | 2 T25357 | hypothetical prote |
| 27 | 391.5 | 15.0 | 445 | 2 T23590 | hypothetical prote |
| 28 | 382.5 | 14.6 | 425 | 2 A90055 | hypothetical prote |
| 29 | 379.5 | 14.5 | 473 | 2 T31717 | hypothetical prote |

| | | | | | |
|----|-------|------|-----|----------|--------------------|
| 30 | 373 | 14.3 | 455 | 1 H69752 | probable glucarat |
| 31 | 371 | 14.2 | 485 | 2 T24115 | hypothetical prote |
| 32 | 369 | 14.1 | 452 | 2 AE0861 | probable glucarate |
| 33 | 368.5 | 14.1 | 478 | 2 T33942 | hypothetical prote |
| 34 | 366.5 | 14.0 | 462 | 2 T34365 | hypothetical prote |
| 35 | 365.5 | 14.0 | 420 | 2 S44900 | ZK52.10 protein - |
| 36 | 361 | 13.8 | 450 | 2 A65061 | probable glucarate |
| 37 | 358.5 | 13.7 | 428 | 2 AD0675 | membrane transport |
| 38 | 357 | 13.6 | 450 | 2 A91085 | probable transport |
| 39 | 357 | 13.6 | 450 | 2 C85930 | probable transport |
| 40 | 346.5 | 13.2 | 455 | 2 T34366 | hypothetical prote |
| 41 | 339.5 | 13.0 | 659 | 2 T33557 | hypothetical prote |
| 42 | 338.5 | 12.9 | 422 | 1 A69853 | hexuronate transpo |
| 43 | 334.5 | 12.8 | 493 | 2 T19383 | hypothetical prote |
| 44 | 331.5 | 12.7 | 516 | 2 T27092 | hypothetical prote |
| 45 | 326 | 12.5 | 487 | 2 T23776 | hypothetical prote |

RESULT 1

G88553

protein C38C10.2 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001

C:Accession: G88553

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio

A:Reference number: A75000; PMID:99069613; PMID:9851916

A>Note: see websites genome.wustl.edu/gsc/C-elegans/ and www.sanger.ac.uk/projects/C-

A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;

A:Accession: G88553

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-493 <STO>

A:Cross-references: GB:chr_III; PIDN:CAA79549.1; PID:g3874873; GSPDB:GN00021; CESP:C

C:Genetics:

A:Gene: C38C10.2

A:Map position: 3

Query Match

Best Local Similarity

Matches

27

8

87

63

147

122

207

182

264

242

320

302

380

37.0%

40.5%

193

Conservative

89

Mismatches

179

Indels

15

Gaps

6

Db 362 VASGYCGGQDVVLIIITTCMAISGQYAGFVVVNYLEIAPPFGTVMGTNTISALAGI 421
QY 439 IGPIIARSLPTENTIGEWQTFECIAAIVNVEGAIFFTFLFAKGEVQNWNA-ISDHQGH 493
Db 422 ISPAVSYLTPNGTQEWQVWLVLTAGILITGALLFSIFASGEVQPWAKUTAEHGH 477

RESULT 2
S28286
hypothetical protein C38C10.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 16-Feb-1997
C:Accession: S28286
R:Thomas, K.
submitted to the EMBL Data Library, December 1992
A:Reference number: S28285
A:Accession: S28286
A:Molecule type: DNA
A:Residues: 1-472 <THO>
A:Cross-references: EMBL:219153
C:Genetics:
A:Introns: 50/3; 287/3; 351/3; 412/3
C:Keywords: transmembrane protein

Query Match 36.3%; Score 950; DB 2; Length 472;
Best Local Similarity 41.0%; Pred. No. 7.6e-69;
Matches 187; Conservative 86; Mismatches 173; Indels 10; Gaps 5;

QY 47 FFGFVLYSLRVNLSVALVMDVSNNTAKDNRTSYECAHSAPIKVLHNTGKKYRWDAE 106
Db 2 FFGCLVTYMMRTNMSFAVVCVMENKTDGTGVEKYSRCGKEMTPVESNSSVIG-EFDWDKQ 60
QY 107 TQGWILGSFFYGYIITQIPGYSVASRSGKLLGLGFGIFATAITFTPLPLAADFGVGA 166
Db 61 TTGVLVSFFYIGTIGSIOIGHLASRGGRRVVFVTILGSALLTLNPNVAARTSEYALAI 120
QY 167 LRALEGGEVGYTPAMHAMSSWAPPLERSKLLSISVAGAQGLTVVSLPSLGVCVY--- 223
Db 121 LRAAIGLQCATFPAMHTMWSVNGPPLLSVLTVGVYAGAQIGNVIVLPSGLCEYGF 180
QY 224 MNVTYVFFYGVIGIWIIFILWICLVSDTPTHTTITPYEKEYILSSLKNOLSSQ----KS 279
Db 181 GMPSPFIYIIGVFGVLTAVWYVSSDKPATPRITPEEKQYIVTAVESAMSGDKDTGKVP 240
QY 280 VPIPMKLSPLWAIVVAHFYSNWTFTLTLLPTTMKEVLRNIOENGFLSAVPYLGW 339
Db 241 TPWIKILTSPAVWACHAGHPAGWGAYTMLVSLPSLKDVLGILNLSLGAVALSIPIYAF 300
QY 340 LCMILSQQAADNLRARNFNTLWVRVFSLIGMIGPAIFLVAAGFTGCDYS-LAVAFLLTI 398
Db 301 LAIAGGVLAADTLRSKGLSTLNRRAAMLVALIGQIFLVASGYCGGQDVLVLIIFITC 360
QY 399 STTLGGFCSSGFSINHLDIAPSVAGILLGTNTFAIPGMIGPIIARSLTPNTIGEWQT 458
Db 361 GMAISGQYAGFVVVNYLEIAPPFGTVMGTNTISALAGISPAVSYLTPNGTQEWQW 420
QY 459 VFCIAAIVNVEGAIFFTFLFAKGEVQNWNA-ISDHQGH 493
Db 421 VLMLTAGILITGALLFSIFASGEVQPWAKUTAEHGH 456

RESULT 3
I59302
brain specific Na+-dependent inorganic phosphate cotransporter - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C:Accession: I59302
R:Ni, B.; Rosteck, P.R.; Nadi, N.S.; Paul, S.M.
Proc. Natl. Acad. Sci. U.S.A. 91, 5607-5611, 1994
A:Title: Cloning and expression of a cDNA encoding a brain-specific Na(+)-dependent ino
A:Reference number: I59302; MUID:94261635
A:Accession: I59302

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-560 <RES>
A:Cross-references: EMBL:U07609; NID:g507414; PIDN:AAAL19646.1; PID:g507415

Query Match 35.4%; Score 927; DB 2; Length 560;
Best Local Similarity 39.4%; Pred. No. 6.6e-67;
Matches 184; Conservative 89; Mismatches 168; Indels 26; Gaps 4;

QY 30 EPAPVCCSA-----RYNLAFSLFFGVLYSLRVNLSVALVMDVSNNTAKDNRTSYECA 84
Db 48 DPPVWDCTCGLPRLRYIIATIMSGLGCISIRGICNLGVAIVSMVNNSTHRRGHVVVQKA 107
QY 85 EHSAPIKVLHNTGKKYRWDAETQGWILGSFFYGYIITQIPGYSVASRSGKLLGLGFI 144
Db 108 Q-----FNMDPETVGLIHGFFWGYIVTQIPGGFICOKFAANRVFGFAIV 152
QY 145 ATAIFTLTPLAADFGVALVLALEGLGEGVYTPAMHAMSSWAPPLERSKLLSISA 204
Db 153 ATSTLNLIPSAARVHYGCVIFVRILQGLVEGYTPACHGINSKWAPPLERSLATATFC 212
QY 205 GAOLGTVWSLPSGVICYNNWTYVFFYGVIGIWIIFILWICLVSDTPTHTTITPYEKE 264
Db 213 GSYAGAVVAMPLAGVLVQYSGSSVFYVYGSFGIFWFLWLVSYESPALHPSISEERK 272
QY 265 YILSSLKNOLSSQKSV-----PWIMPLKSLPLWAIVVAHFYSNWTFTLTLLPTMYKEV 319
Db 273 YIEDALGESAKLMNPYTKFNTPMWRREFTSMPYAILVANFCRSWTFYLLLSQPATFEV 332
QY 320 LRNIQENGFSLSAVYLGWCLMILSQQAADNLRARNFNTLWVRVFSLIGMIGPAIFL 379
Db 333 FGEIISKVGLVSALPHLVMTIIVPIGQIADFLSRHIMSTTNVRKLMNCGGFGMEATLL 392
QY 380 VAAGFTGCDYSLAVAFLLTSTTLGGFCSSGFSINHLDIAPSVAGILLGTNTPTATPGMI 439
Db 393 LVVGY-SHSGKGAISFLVLAVGFSFGAISGFNVNHLDIAPRYASILMGISNGVGTLSGW 451
QY 440 GPIIARSLPTENTIGEWQTFECIAAIVNVEGAIFFTFLFAKGEVQNWNA 486
Db 452 CPIIVGAMTKHKTREEWQYVFLIASLVHYGGVIFYGVFASGEKQPWA 498

RESULT 4
T43650
probable sodium-dependent inorganic phosphate cotransporter - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: T43650
R:Lee, R.Y.N.; Sawin, E.R.; Chalfie, M.; Horvitz, H.R.; Avery, L.
submitted to the EMBL Data Library, September 1998
A:Description: EAT-4, a homolog of a mammalian sodium-dependent inorganic phosphate c
A:Reference number: Z22599
A:Accession: T43650
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-563 <LEE>
A:Cross-references: EMBL:AF095787; PIDN:AAC64972.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: eat-4
A:Map position: 3
C:Function:
A:Description: is necessary for glutamatergic neurotransmission

Query Match 31.3%; Score 819; DB 2; Length 563;
Best Local Similarity 36.3%; Pred. No. 3.4e-58;
Matches 165; Conservative 88; Mismatches 178; Indels 24; Gaps 6;

QY 36 CSARYNLAFSLFFGVLYSLRVNLSVALVMDVSNNTAKDNRTSYECAHSAPIKVLIN 95
Db 53 CRKRWLLAILANMGFMISFGIRCNGCAKTHMYKNYTD-----PYGKVHM 97

| | | | |
|---------------------|-----|---|-----|
| Qy | 96 | QTGKKYRWDAETGOWILGSGFFGYIITQIPGGVYASRSKGKLLGFGIFATAIFTLTPL | 155 |
| Db | 98 | H--EFNWTIDELSVMESSYFYGYLVTQIPAGFLAAKFPFNKLFGRGCVGAGFLNILLPY | 154 |
| Qy | 156 | AADFGVGVCAVA-LRALEGLGEGVYTPAMHAMSSWAPPLERSKLLSISVAGAQLGTVVSL | 214 |
| Db | 155 | GFKVSDYLVAFTQITGVLGVGVCPAMGVGWRYWAPPMERSKLATATGTSYAGAVGL | 214 |
| Qy | 215 | PLSGVICYNNWTVYFFFGIVGIWFIWILMICIVSDTPETHKITTPYEKEYI--LSSLK | 271 |
| Db | 215 | PLSAFLSVYSMAAPFYLVGCVGIWAILWFCVTFEKPFAHPTISOEEKIFIEDAIGHVS | 274 |
| Qy | 272 | NQLSSOKSPWIPMLKSLPLWALVVAHYSNNTFYTLPLLPPTYMKVEYRFRNIQNGFLS | 331 |
| Db | 275 | NTHPTIRSPKAIATVSKPWAIIIVANFARSFTYLLQLNQLTKYKCALGMKIASGLLIA | 334 |
| Qy | 332 | AVPYLGCWLCMILSGQAADNLRAWNFSITLWVRVPSLIGMTGPAIFLVAAGFIGCDYSL | 391 |
| Db | 335 | AIPHLVMGCVVLMMGGQLADYLRNKLSTTAVNRKIPNCGGFGGEAFMLIVATYTSDD-TT | 393 |
| Qy | 392 | AVAFLTITSTLGGFCSSGFSINHLDIAPSVAGILGTTTFTATIPGMIGPIIARSUTPEN | 451 |
| Db | 394 | AIMALLAAVCKSGFAISGFNVNHLDTAPRYAILMGFSNGIGTLAGLTCPFTVEAFTAH | 453 |
| Qy | 452 | TIGEWOTVFCIAAIVFGAIFTLTFAKEGVQNW 486 | |
| Db | 454 | KHG-WTSVFLASLIHFTGVTFFVAVASGELQEW 487 | |
| RESULT | 5 | | |
| H88548 | | protein ZK512.6 [imported] - Caenorhabditis elegans | |
| C:Species: | | Caenorhabditis elegans | |
| C:Date: | | 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2000 | |
| C:Accession: | | H88548 | |
| R:anonymous, | | The C. elegans Sequencing Consortium. | |
| Science | | 282, 2012-2018, 1998 | |
| A:Title: | | Genome sequence of the nematode C. elegans: a platform for invest. | |
| A:Reference | | number: A75000; PMID:99069613; PMID:9851916 | |
| A:Note: | | see websites genome.wustli.edu/gsc/C_elegans/ and www.sanger.ac.uk/ | |
| A:Note: | | published errata appeared in Science 283, 35, 1999; Science 283, 2 | |
| A:Accession: | | H88548 | |
| A:Molecule | | type: DNA | |
| A:Residues: | | 1-576 <STO> | |
| A:CROSS-references: | | GB:chr_III; PIDN:CRAA80150.1; PID:g3881690; GSPDB:GN000 | |
| C:Genetics: | | | |
| A:Gene: | | ZK512.6 | |
| A:Map position: | | 3 | |

Db 288 NTHPTKSIIPWKAIVTSKPVWALIVANFARSWTFYLLDQLNQTMYKEAGMKIADSGLLA 347
 Qy 332 AVPYLGCWLMILSGQAADNLRARWPNFSTLWVRVFSLLIGMIGPAIFLVAAGFIGDYSL 391
 Db 348 AIPLHYMGCVVLGMQDLADYLRNKKILSTTAVRKIFNCGGFGGEAAFMILVAVYTTSD-TT 406
 Qy 392 AVAFLLTSTTLGGFCSSGFSINHLDTAPSTAGILLGTITWTFTATIPGMIGPIIARSLTPEN 451
 Db 407 AIMALIAAVGMSGFATSGFNVNHLDTAPRYAAILMGFSNGIGTLAGLTCPFVTEAFTAH 466
 Qy 452 TIGEWQTVFCIAAIVNFGAIFFTLFAKGEVQNW 486
 Db 467 KHG-WTSVFLIASLIHFTGVTFAVYASGELQEW 500

 RESULT 6
 A56410
 sodium/phosphate transport protein, renal - rabbit
 N:Alternate names: sodium/phosphate cotransporter
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 05-Nov-1999
 C:Accession: A56410; S27951
 R:Werner, A.; Moore, M.L.; Mantel, N.; Biber, J.; Semenza, G.; Murer, H.
 Proc. Natl. Acad. Sci. U.S.A. 88, 9608-9612, 1991
 A:Title: Cloning and expression of cDNA for a Na/P-i cotransport system of kidney cor
 A:Reference number: A56410; MUID:92052140
 A:Accession: A56410
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-465 <WER>
 A:Cross-references: GB:M76466; NID:gl65689; PIDN:AAA31461.1; PID:gl65690
 C:Keywords: kidney; transmembrane protein

 Query Match 30.9%; Score 809.5; DB 2; Length 465;
 Best Local Similarity 36.0%; Pred. No. 1.6e-57;
 Matches 164; Conservative 86; Mismatches 196; Indels 9; Gaps 2;

 Qy 33 PVCCSARYNLAFISFGFFWLYSLRVNLVALVDVDSNTAKDNRTSYSCABHSAPIKV 92
 Db 11 PCFCFRYVLALEHFHFCNIVIIAQRMCMLSTWAMV-----NNTNLHGSPNTSAER 62
 Qy 93 LHNQTGKKYRWDAAETOGWILGSFPFYIIITQPGYVVASRGGKLLGLGFGIFATAIFTLF 152
 Db 63 LDNTKPNVYNWSPDVQGLIFSSIFYGAFLIQIPGVYISGYISKKLIGLGFALFUSLSVIF 122
 Qy 153 TPLAADFGVALRALALEGLGEGVTVPAMHAMSSWAPPLERSKLLSISYAGAQLGTVV 212
 Db 123 IPQAAAGETWIIICVRVVGITQGTVTTAQHEIWKWAPPLERGLTSMLSGFLGPF 182
 Qy 213 SLPLSGVICYMMNWTYVYFPGIVGIIWFIWLICLVSDTPETHKTITPYKKEYILSSLN 272
 Db 183 VLLVTGIIICESLGPWMPYIFYGACGACVLLIVFLVYDDPKDHPCVSLHEKEYITSSLIQ 242
 Qy 273 QLSS-QKSVPIWPKLSLPLWAIWVAHFSYNWTFYLLTLPTYMKEVLRFNIOENGFLS 331
 Db 243 QGSSTROSLPTKAMIKSLPLWAISSFCFAYLWTVSRIVTYPTLLINSMHLVDIRENGLS 302
 Qy 332 AVPYLGCWLMILSGQAADNLRARWPNFSTLWVRVFSLLIGMIGPAIFLVAAGFIGCDYSL 391
 Db 303 SLPLFAWICGVIAGHTADFLMSRNLSLTAIRKLFTAIGLLLPVFSMCLLYLSGFGYS 362
 Qy 392 AVAFLLTSTTLGGFCSSGFSINHLDTAPSTAGILLGTITWTFTATIPGMIGPIIARSLTPEN 451
 Db 363 TITELIANANSSSPCLGALINALDLAPRYVYFTKGVYTTLLIGMTGGMTSSTVAGLFLSQD 422
 Qy 452 TIGEWQTVFCIAAIVNFGAIFFTLFAKGEVQNW 486
 Db 423 PESSWFKIFLUMSIINVISVIFLIIFAKAEIQDW 457

S69915

sodium-phosphate transport system 1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999
C:Accession: S69915
R:Chong, S.S.; Kozak, C.A.; Liu, L.; Kristjansson, K.; Dunn, S.T.; Bourdeau, J.E.; Hughes, A.M. J. Physiol. 268, 1038-1045, 1995
A:Title: Cloning, genetic mapping, and expression analysis of a mouse renal sodium-dependent anion channel protein
A:Reference number: S69915
A:Accession: S69915
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-465 <CHO>
A:Cross-references: EMBL:X77241; NID:g887522; PIDN:CAA54459.1; PID:g887523

Query Match 30.7%; Score 802.5; DB 2; Length 465;
Best Local Similarity 36.5%; Pred. No. 5.8e-57;
Matches 166; Conservative 89; Mismatches 191; Indels 9; Gaps 2;
QY 33 PVCCSARYNLAFSLSEFGFFVLYSLRVNLVALVDMVDSNTTAKDNRTSYECAHSAPIKV 92
DB 11 PGCFSTRYGLAILLHFCNTAINMAQRCLNLTVMAMV-----NNTGSPHLSNESVEM 62
QY 93 LHNQTKKYRWDAETQGWILGSPFYGIITQIPGGYVASRSRGKLLGLGFIATATITL 152
DB 63 LDNVKNPVYSWSPDIOGLILSSVFFGCMVVVQAPVGLSGIYPMKRIIGSSLSLSLSLL 122
QY 153 TPLAADFGVALVALRALGEGVITYPAMHAMSSWAPPLERSKLLSYAGAGLGTVV 212
DB 123 IPPAAQVGAALVTCRVLOIAQGVSTGQHEIWKWAPPLERGRUITSLSGFMVGPFI 182
QY 213 SLPLSGVICYNNWTVYVFFGVLGIWILWICLVSDTPETHKTIPTPYEKEYILSLKN 272
DB 183 VLVSGFICDLGWPVVFIFGVLGVLSSWFFLEFDKDPKHPYSSSEKDYIIISLMQ 242
QY 273 QLSS-QKSPWPIMPKSLPLMAIWAHFSYNWTFYLLTLLPTMYKEVLRNFNIQENGFLS 331
DB 243 QAASSGROSLPIKAMKSLPLWAILLSFAFINSLSLVYTPFTFISTVLHVNVRENGLS 302
QY 332 AVPLICWCILMSLQGAADNLARNWFLWVRVSLGMDGPAIFLVAAGFIGDYSL 391
DB 303 SLPYLLAYICGLLAGOMSDFFLTKRIFSIVTKRLETLTGSCFVPVIFMCLLYLSNFYS 362
QY 392 AVAFLITSTLGGSCGFSINHLDIAPSAGVILLGITNTFATIPGMIGPIIARSLTPEN 451
DB 363 TVIFLTLANSTLSFSVCGQLINALDIPRYGFLKAVTALIGMFGGLISSTLAGLILNOD 422
QY 452 TIGEWTVFCIAAANVFGAIFFTLFAKGEVQNW 486
DB 423 PEYAWHKISFLMAGINVTCLVFFLFAKGEIQDW 457

RESULT 8
23589
hypothetical protein K10G9.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T23589; T24636
R:Wortimore, B.
submitted to the EMBL Data Library, August 1994
A:Reference number: Z19765
A:Accession: T23589
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-573 <WIL>
A:Cross-references: EMBL:Z36282; PIDN:CAA85289.1; GSPDB:GN00021; CESP:K10G9.1
A:Experimental source: clone K10G9
R:Buck, D.
submitted to the EMBL Data Library, February 1995
A:Reference number: Z19915
A:Accession: T24636
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-573 <WIL>
A:Cross-references: EMBL:Z48055; PIDN:CAA88135.1; GSPDB:GN00021; CESP:K10G9.1
A:Experimental source: clone T07A5
C:Genetics:
A:Gene: CESP:K10G9.1
A:Map position: 3
A:Introns: 38/3; 87/3; 224/1; 272/3; 322/3; 444/3; 476/3; 496/1; 517/2

Query Match 30.3%; Score 792.5; DB 2; Length 573;
Best Local Similarity 37.1%; Pred. No. 4.7e-56;
Matches 171; Conservative 85; Mismatches 164; Indels 41; Gaps 9;
QY 39 RYNLAFLSFFGFFVLYSLRVNLVALVDMVDSNTTAKDNRTSYECAHSAPIKVLINQGT 98
DB 47 RMOIALLAHFGFAISFGIRSNFGVAKNRMINFTDAYGE-----VHE--- 88
QY 99 KKYRWDAETQGWILGSPFYGIITQIPGGYVASRSRGKLLGLGFIATATITL 158
DB 89 KEFFWTGTEVGMESSEFFGYAASQIPAGVIAAKFAPNKLFLMLGILFASLLNIVTAICLN 148
QY 159 FGVA---LVALRALGEGVITYPAMHAMSSWAPPLERSKLLSYAGAGLGTVVSLP 215
DB 149 FHPFTDIFVMVIVQMQLGALVCYPAMHGVMKYWAPPLERSKLATTTFTGASGVVMGLP 208
QY 216 LSGVICYNNWTVYVFFGVLGIWILWICLVSDTPETHKTIPTPYEKEYILS-----SL 270
DB 209 ASAYLSYHSWSWTFYVFGALGIWLSILWYVSGTSPETHGYISADEKKYITEKVSVA 268
QY 271 KNLSSOKSVPTPMLKSLPLMAIWAHFSYNWTFYLLTLLPTMYKEVLRNFNIQENGFL 330
DB 269 KNM--TLTILPWRDMTSTAVWAIICSCRSWSFLLGNQLTYMKDVLHDIKNSGLI 326
QY 331 SAVPYLGWCILMSLQGAADNLARNWFLWVRVSLGMDGPAIFLVAAGFIGC--- 387
DB 327 AIFPQLMGCVITLTSQGLSYLRSGKMSKMSKMSKMSKMSKMSKMSKMSKMSKMSK 380
QY 388 --DYSLAVAFITISTLGGSCGFSINHLDIAPSAGVILLGITNTFATIPGMIGPIIAR 445
DB 381 VRDPVIAVTFLLIACSGAGAVLSGFNVNHFDAIRHAPILMGIANGLAGIAG--VGGIVTN 439
QY 446 SLTPNTIGEWTVFCIAAANVFGAIFFTLFAKGEVQNW 486
DB 440 SLTYQNPDC--WQVFLANSLIDIFGIIFLFAKGDVLPWA 479

RESULT 9
A48916
sodium phosphate transport protein 1, renal - human
C:Species: Homo sapiens (man)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Nov-1999
C:Accession: A48916
R:Chong, S.S.; Kristjansson, K.; Zoghbi, H.Y.; Hughes, M.R.
Genomics 18, 355-359, 1993
A:Title: Molecular cloning of the cDNA encoding a human renal sodium phosphate transporter
A:Reference number: A48916; MUID:94117004
A:Accession: A48916
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-467 <CHO>
A:Cross-references: GB:X71355; NID:g450531; PIDN:CAA50490.1; PID:g450532
C:Genetics:
A:Gene: GDB:SLC17A1; NPT1
A:Cross-references: GDB:I41889; OMIM:182308
A:Map position: 6p23-6p21.3
Query Match 30.2%; Score 791.5; DB 2; Length 467;
Best Local Similarity 37.4%; Pred. No. 4.5e-56;
Matches 180; Conservative 79; Mismatches 183; Indels 39; Gaps 7;
QY 23 LQRAPRAEP--APVCCSARYNLAFSLSEFGFFVLYSLRVNLVALVDMVDS-----NTT 73

Db 1 MDMNRLLPKKPGCSFRIGLSELVHCCNVITTAQRACLNLTVMVWVNSTDPHGLPNTS 60
QY 74 AKDNRTSYECASHPAPIKVLHNOTGKRYWDAETOCGILGSGFFGYIITQIPGGYVASRS 133
Db 61 TK-----KLLDNKPMYNWSPDIOGIISSYSYGVIIIOVPVGFSGIY 105
QY 134 GKLLGFGIFATAITFTPLAADFGVALVRALEGLGEGVTYPAMHAMSSWAPPL 193
Db 106 STKKMIGFALCLSSLLIPPAAGIGVAVVVCRAVQAAQIVATAQFEIYVWAPPL 165
QY 194 ERSKLLTSYAGAOGLTVVSLPLSGVICYYMNTVYVFFGIVGIIWFIWILICLVSDTPE 253
Db 166 EGRUTSMTSGFLGPPFTLVLTGVICESLGMVYIFGACGCAVCLLWFVLYDDPK 225
QY 254 THKTTIPYEKILSKNOL--SSOKSVPIWMLKSLPLWAIWVAHFSYNWTFYLLTLL 312
Db 226 DHPICISIEKEVITSLVQVSSRSQSLPIKAILKSLPVAISISGFTFFWSHNMTLYT 285
QY 313 PTYMKVLRNFNIQENGLFSAVPLGCLWMLSGOAGNLRARNFSTLWVRVFSLIGM 372
Db 286 PMFINSMLHVNKENGFLSSLPYFAWICGNLAGLSDFELTRNLSVIAVRKLFATAAGF 345
QY 373 ICPAIFLVAAGFIGDYSLAVAFITSTTLGFCSSGSINHLDTAPSYAGILLGITNTF 432
Db 346 LPAIFGVCLPYLSFTYSIVIFILAGATGSCFCLGGVFINGLDIAPRYFGFI----KAC 401
QY 433 ATIPGMIGIARSLT-----PENTIGEWOTVFCIAAANVFGAIFFTLFAKGEVQNW 485
Db 402 STLTGMIGLIASTLTGLLKQDPESA---WFKTFILMAAINVTGLIFLIVATAEIQDW 458
QY 486 A 486
Db 459 A 459
RESULT 10
I39473
Na+-dependent phosphate cotransporter - human
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Nov-1999
C:Accession: I39473
R: Miyamoto, K.; Tatsumi, S.; Sonoda, T.; Yamamoto, H.; Minami, Y.; Takeda, Y.; Biochem. J. 305, 81-85, 1995
A: Title: Cloning and functional expression of a Na(+)-dependent phosphate co-transporter
A: Reference number: I39473; MUID: 95126933
A: Accession: I39473
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: mRNA
A: Residues: 1-465 <RES>
A: Cross-references: GB:D28532; NID:g639841; PIDN:BAA05888.1; PID:g639842

Query Match 29.6%; Score 775.5; DB 2; Length 465;
Best Local Similarity 37.3%; Pred. No. 8.7e-55;
Matches 175; Conservative 79; Mismatches 178; Indels 37; Gaps 6;
QY 33 PVCCSARYNLAFSPFGFVLYSLRVNLSVALVDMVDS-----NTAKDNRTSYECAE 85
Db 11 PGFCFRGLSFLVHCNCVITTAQMLCLNLTVMVWVNSTDPHGLPNTSK----- 60
QY 86 HSAPIKVLHNOTGKRYWDAETOCGILGSGFFGYIITQIPGGYVASRSGLLLGFGIPA 145
Db 61 -----KLLDNKPMYNWSPDQIILSSYSYGVIIIOVPVGFSGIYTKKMGIFALCL 115
QY 146 TAIFTLPTPLAADFGVALVRALEGLGEGVTYPAMHAMSSWAPPLSKLLSISVAG 205
Db 116 SSVLSLIPPAAGIGVAVVVCRAVQAAQIVATAQFEIYVWAPPLGRLTSMSTSG 175
QY 206 AOLGTWVSLPLSGVICYYMNTVYVFFGIVGIIWFIWILICLVSDTPEHKTITPYEKY 265
Db 176 FLTGPPFVLLVTGVICESLGMVYIFGACCCAVCLLWFVLYDDPKDHPICISIEKEBY 235

QY 266 ILSSLKNOI--SSOKSVPIWMLKSLPLWAIWVAHFSYNWTFYLLTLLPTYMKVLRNFI 324
Db 236 ITSSLVQVSSRSQSLPIKAILKSLPVAISISGFTFFWSHNMTLYTDMFINSMLHVN 295
QY 325 OENGLFSAVPLGCLWMLSGOAGNLRARNFSTLWVRVFSLIGMIGPAIFLVAAGF 384
Db 296 KENGLFSSLPYFAWICGNLAGLSDFELTRNLSVIAVRKLFATAAGFLLPAIFGVCLPY 355
QY 385 IGCDSLAVAFITSTTLGFCSSGSINHLDTAPSYAGILLGITNTFTATPGMIGPIIA 444
Db 356 LSTFYSIVIFILLAGATGSCFCLGGVFINGLDIAPRYFGFI----KACSTLTGMIGLIA 411
QY 445 RSLT-----PENTIGEWOTVFCIAAANVFGAIFFTLFAKGEVQNW 486
Db 412 STLTGLILKQDPESA---WFKTFILMAAINVTGLIFLIVATAEIQDW 457
RESULT 11
T24633
hypotheetical protein T07A5.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24633
R: Buck, D.
Submitted to the EMBL Data Library, February 1995
A: Reference number: Z19915
A: Accession: T24633
A: Status: preliminary; translated from GB/EMBL/DBJ
A: Molecule type: DNA
A: Residues: 1-544 <WIL>
A: Cross-references: EMBL:Z48055; PIDN:CAA88134.1; GSPDB:GN00021; CESP:T07A5.3
A: Experimental source: clone T07A5
C: Genetics:
A: Gene: CESP:T07A5.3
A: Map position: 3
A: Introns: 38/3; 87/3; 272/3; 322/3; 444/3; 476/3; 496/1; 517/2
Query Match 29.3%; Score 766.5; DB 2; Length 544;
Best Local Similarity 35.6%; Pred. No. 5.5e-54;
Matches 164; Conservative 90; Mismatches 166; Indels 41; Gaps 9;
QY 39 RYNLAFSPFGFVLYSLRVNLSVALVDMVDSNTTAKDNRTSYECASHPAPIKVLHNOTG 98
Db 47 RQOITAILAHFGPAISGIRSNFEGVAKRMVNNFTDAYGE-----VHE--- 88
QY 99 KKYRWDAAETOGWILGSGFFGYIITQIPGGYVASRSGLLLGFGIPATAIFTLFTPLAAD 158
Db 89 REFLWTGAEVGMWMESSFFGYAASQIPAGVLAAKFAPNKFIMLGILVASFMNLSAISFN 148
QY 159 FGVA---LVALRALGELGEGVTYPAMHAMSSWAPPLSKLLSISYAGAOGLTVVSLP 215
Db 149 FHPYTDIFVWVQAVQGLALGVLYPAMHGVKFWAPPLERSKLATTAFTFGSSVGVMTGLP 208
QY 216 LSGVICYYMNTVYVFFGIVGIIWFIWILICLVSDTPEHKTITPYEKYILS-----SL 270
Db 209 ASAYLSHSFSWSPFYVFGVGIWLSLIWVYSSHPETHGYISDDEKKQVTEKIGDVAV 268
QY 271 KQLSOKSVPIWMLKSLPLWAIWVAHFSYNWTFYLLTLLPTYMKVLRNFNIQENGL 330
Db 269 KNN--SLTTLPRMDMTSSAVWAIITICTFCRSGWFFLLGNLTQYMKVDLHIDIKNSGFI 326
QY 331 SAVPYLGCWCLMILSGOAGNLRARNFSTLWVRVFSLIGMIGPAIFLVAAGFICG--- 387
Db 327 SIFPQGMCIIVLATQGLCDYLRSSCKMSTEAVRKSNTFTG-----FIVEAMMLCLAF 380
QY 388 --DYSIAVAFITSTTLGFCSSGSINHLDTAPSYAGILLGITNTFTATPGMIGPIIA 445
Db 381 VRDPVIAVTCVITACTGSGVSLSGFNHFDIAPRYATILMGLIANGLAGVAG-VGGMVTN 439
QY 446 SLTPENTIGEWOTVFCIAAANVFGAIFFTLFAKGEVQNW 486
Db 440 TVTYQNDPQD-KWVFLAMAIIDIFGVIFLIFAKGDVLPWA 479

RESULT 12
H84698
hypothetical protein At2g29650 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84698
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: H84698
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-512 <STO>
A:Cross-references: GB:AE002093; NID:g3582333; PIDN:AAC35230.1; GSPDB:GN00139
A:Gene: At2g29650
A:Map position: 2

Query Match 28.4%; Score 744; DB 2; Length 512;
Best Local Similarity 33.9%; Pred. No. 3.3e-52;
Matches 166; Conservative 83; Mismatches 187; Indels 54; Gaps 9;

QY 2 KSPVSDLPADSGEGSDRTPLLQAPRA-----EPAPVCCSARYNLAFLSPFGFVLVS 55
DB 63 KSESSSYVEGDKVGNNDVSDSPSSVLPWMEFP-----KRWIVILCFSAFLCNM 118
QY 56 LRYNLSVALVDMVDSNTAKDNRTSYECAHSAPIKVLHNQTKGKYRWDATOGWILGSF 115
DB 119 DRYNMSIALP-----SAEYGNPATVGLIQSSF 148
QY 116 FYGIITQIPGGVVASRSGKLLGRGIFATAFTFTPLAADFGVGLVALRALEGLGE 175
DB 149 FWGYLLTQIAGGIWADTVGKRVGLFGVIWSTATILTPVAAKLGPLVLLVRAEFGVGE 208
QY 176 GVTPYPAHAWMSWAPPLERSKLLSISYAGAQIGTVVSLPSLVGICVYNNWTVVFFGI 235
DB 209 GVAMPAMNNILSKWVPQKRSRLALVSGMYLGSVTVGLAFSPFLIHQFGWPSVFFSGS 268
QY 236 VGIWFLWILCLVSDRTPETHKTTTPYEKEYILLSSKNQLSOKSVPIWPKLSPLWAI 295
DB 269 LGIVWLTLWTKAESPLEDPILLPERKLIADNCASK-EPYKSIPIRWLLSKPPVWALI 327
QY 296 VAHFSYNTWTFYLLTLPTMYKEVLRNIOENGFLSAVPLYGCWLCLMSGQA-----ADN 351
DB 328 SCHFCNMGTFILLTWMTYYHOVLKFNLMESGLLSVFP-----WMTMAISANAGGWIADT 383
QY 352 LRARNFSLWVRVSLGIMGPAIFLVAAGFIGDCYSLAVAFLLISTLGGFCSSGFS 411
DB 384 LVSR-GFSTVNRKIMQITGFLGPAFLLQLKHIDSP-TMVLVCMACSQGTDAFSQSGLY 441
QY 412 INHLDIAPSVAGILLGITNTFATIPGMIGPIIARSITPENTIGEMOTVFCIAAINVEGA 471
DB 442 SNHQDIAPRYSGVLLGSNTAGVLGTAATGCHILQH---GSWDDVTISVGLVLVGT 498
QY 472 IFFTLFAKGE 481
DB 499 VIWNLFSTGE 508

RESULT 13
B89135
protein F25G6.7 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: B89135
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: B89135
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-501 <STO>
A:Cross-references: GB:chr_V; PIDN:AAC25800.1; PID:g2384837; GSPDB:GN00023; CESP:F25G
C:Genetics:
A:Gene: F25G6.7
A:Map position: 5

Query Match 22.4%; Score 585; DB 2; Length 501;
Best Local Similarity 29.9%; Pred. No. 2.1e-39;
Matches 149; Conservative 101; Mismatches 201; Indels 48; Gaps 13;

QY 4 PVSDDLAPSDGEGSDRTPLLQOR-APRAEPAPVCCSARYNLAFLSFFGFVLYSLRVNLSV 62
DB 6 PISDGIP-----SQQRPLTYRWP-----SWREMTSIMLCFCFCGCVHLMNSMGM 50
QY 63 ALVDMVDSNTAKDNRTSYECAHSAPIKVLHNQTKGKYRWDATOGWILGSFFFYGIIT 122
DB 51 AIYCMVNSSAT-YDNET---YPENAAPL-----LDWSSDEQGYIFSFAFNAGLLVM 96
QY 123 QIPGGVVASRSGK--LLLGFGIFATAFTFTPLAADFGVGLVALRALEGLGEGVTYP 180
DB 97 LFTGG-MADKFNKAKYMLVSLVAISLANFTL--PWAPISVYWAIFSRFLVGFADALLQP 153
QY 181 AMHAMSSWAPPLERSKLLSISYAGAQIGTVVSLPSGVIC----YMMNWTYVYFFFGIV 236
DB 154 AMNSLITRWFPTSESRVALGATGGGIGTIIIPAGALCSQTEIFGGWPSFIFLSGFI 213
QY 237 GIWFLWILCLVSDTETHKTTTPYEKEYILLSSKNQLS-----SOKSVPIWPKLSLP 290
DB 214 GVLFIFSYFLGADKPSKQSCISDNELKFI--TISQSEVDGKKRTERKYPWKNILKSGA 271
QY 291 LWAIVVAHFSYNWTFYLLTLPTMYKEVLRNIOENGFLSAVPLYGCWLCLMSLQQAAD 350
DB 272 WNASVISLVCHEPPLMTLWFLPSYLDVHHYSTENGILSALPTVSLWFAKIGSSYLNT 331
QY 351 NLRARNFSLWVRVSLGIMGPAIFLVAAGFIGDCYS-LAVAFLLISTLGGFCSSG 409
DB 332 WLQKNTTWKDKTICKVLNSIGSIGLGVFLAATFLDNEHAMWAVFLCLSMASAGLHTPG 391
QY 410 FSNHLDIAPSVAGILLGITNTFATIPGMIGPIIARSITPENTIGEMOTVFCIAAINVF 469
DB 392 COLALYSVAPAYSGAVTGTFFVFVAVSGIHPITIKMIVKDRTAAEWNLVFIYSTVIAIF 451
QY 470 GAIFFTLFAKGEVONWAS 488
DB 452 PIIFNVWGSTEVOWAKS 470

RESULT 14
T01534
hypothetical protein A_IG005110.nn - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 17-Mar-2000
C:Accession: T01534
R:Andrews, S.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of A. thaliana IG005110.
A:Reference number: Z14347
A:Accession: T01534
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-413 <AND>
A:Cross-references: EMBL:AF013293; NID:g2252823; PID:g2252847
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4

```
Query Match          21.0%; Score 550.5; DB 2; Length 537;
Best Local Similarity 32.4%; Pred. No. 1.4e-36;
Matches 126; Conservative 75; Mismatches 155; Indels 33; Gaps 8;
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 08:25:26 ; Search time 17.92 seconds
(without alignments)
1069.541 Million cell updates/sec

Title: US-09-776-865-4

Perfect score: 2617

Sequence: 1 MKSPVSDLAPSGEGSDRT.....LFAKGEVQNAISDHQHRN 495

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|------------|--------------------|
| 1 | 968.5 | 37.0 | 493 | YLD2_CABEL | Q03567 caenorhabdi |
| 2 | 819 | 31.3 | 576 | YQ06_CABEL | P34644 caenorhabdi |
| 3 | 809.5 | 30.9 | 465 | NPT1_RABIT | Q28722 o renal sod |
| 4 | 802.5 | 30.7 | 465 | NPT1_MOUSE | Q61983 mus musculu |
| 5 | 797.5 | 30.5 | 465 | NPT1_RAT | Q62795 rattus norv |
| 6 | 790.5 | 30.2 | 465 | NPT1_HUMAN | Q14916 homo sapien |
| 7 | 766.5 | 29.3 | 544 | YRT3_CABEL | Q10046 caenorhabdi |
| 8 | 672 | 25.7 | 436 | NPT3_HUMAN | O00624 homo sapien |
| 9 | 603.5 | 23.1 | 401 | NPT4_HUMAN | O00476 homo sapien |
| 10 | 489 | 18.7 | 568 | YKH4_CABEL | P34272 caenorhabdi |
| 11 | 373 | 14.3 | 455 | GUDP_BACSU | P42237 bacillus su |
| 12 | 361 | 13.8 | 450 | GUDP_ECOLI | Q46916 escherichia |
| 13 | 338.5 | 12.9 | 422 | EXUT_BACSU | O34456 bacillus su |
| 14 | 320.5 | 12.2 | 430 | DGOT_ECOLI | P31457 escherichia |
| 15 | 314 | 12.0 | 435 | YFBO_BACSU | P37489 bacillus su |
| 16 | 300 | 11.5 | 444 | GARP_ECOLI | P42613 escherichia |
| 17 | 291 | 11.1 | 453 | XJL1_ECOLI | P39398 escherichia |
| 18 | 281.5 | 10.8 | 472 | EXUT_ECOLI | P42609 escherichia |
| 19 | 261.5 | 10.0 | 456 | GUDP_PSEPU | P42205 pseudomonas |
| 20 | 246 | 9.4 | 429 | YFAP_ECOLI | P76470 escherichia |
| 21 | 226 | 8.6 | 433 | TUB4_AGRVI | Q44470 agrobacteri |
| 22 | 226 | 8.6 | 449 | TUB3_AGRVI | P70786 agrobacteri |
| 23 | 201.5 | 7.7 | 444 | GLPT_BACSU | P37948 bacillus su |
| 24 | 199 | 7.6 | 407 | YB04_HAEIN | P71369 haemophilus |
| 25 | 199 | 7.6 | 498 | YAO1_SCHPO | Q10097 schizosacch |
| 26 | 195.5 | 7.5 | 452 | GLPT_ECOLI | P48194 escherichia |
| 27 | 190 | 7.3 | 451 | PHL1_PSEPU | Q05181 pseudomonas |
| 28 | 189.5 | 7.2 | 480 | GLPT_HAEIN | P96335 haemophilus |
| 29 | 188 | 7.2 | 533 | GLPT_HUMAN | P57057 homo sapien |
| 30 | 187.5 | 7.2 | 413 | MUCK_ACICA | P94131 acinetobact |
| 31 | 184.5 | 7.1 | 403 | MHPT_ECOLI | P77589 escherichia |
| 32 | 174.5 | 6.7 | 427 | YDFJ_ECOLI | P77228 escherichia |
| 33 | 169 | 6.5 | 496 | NANT_ECOLI | P41036 escherichia |

RESULT 1

| ID | YLD2_CABEL | STANDARD; | PRT; | 493 AA. |
|----|--|-----------|------|------------|
| AC | Q03567; | | | |
| DT | 01-FEB-1994 (Rel. 28, Created) | | | |
| DT | 01-NOV-1997 (Rel. 35, Last sequence update) | | | |
| DT | 01-MAR-2002 (Rel. 41, Last annotation update) | | | |
| DE | Hypothetical 52.7 kDa protein C38C10.2 in chromosome III. | | | |
| GN | C38C10.2. | | | |
| OS | Caenorhabditis elegans. | | | |
| OC | Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; | | | |
| OC | Rhabditidae; Peloderinae; Caenorhabditis. | | | |
| OX | NCBI_TaxID=6239; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=BRISTOL N2; | | | |
| RX | MEDLINE=94150718; PubMed=7906398; | | | |
| RA | Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coultson A., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A., Craxton M., Dear S., Du Z., Durbin R., Favello A., Hillier L., Jier M., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., Jones M., Kershaw J., Kirsten J., Laister N., Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R., Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R., Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.; | | | |
| RT | "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans"; | | | |
| RL | Nature 368:32-38(1994). | | | |
| RN | [2] | | | |
| RP | REVISTONS. | | | |
| RA | Jones S.J.M.; | | | |
| RC | Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases. | | | |
| RL | -I- SUBCELLULAR LOCATION: Integral membrane protein (Potential). | | | |
| CC | -I- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1. | | | |
| CC | ----- | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | | | |
| CC | ----- | | | |
| DR | EMBL; Z19153; CAA79549.1; - | | | |
| DR | PIR; S28286; S28286. | | | |
| DR | WormPep; C38C10.2; CE08647. | | | |
| DR | InterPro; IPR003662; sub-transporter. | | | |
| DR | Pfam; PF00083; sugar tr; 1. | | | |
| KW | Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport; | | | |
| KW | Sodium transport. | | | |
| FT | TRANSMEM 10 30 | | | POTENTIAL. |
| FT | TRANSMEM 85 105 | | | POTENTIAL. |
| FT | TRANSMEM 112 132 | | | POTENTIAL. |

| | | | | | | |
|----|-------|-----|-----|---|------------|--------------------|
| 34 | 166 | 6.3 | 412 | 1 | YWFA_BACSU | P39637 bacillus su |
| 35 | 163 | 6.2 | 457 | 1 | PCAK_ACICA | Q43975 acinetobact |
| 36 | 161.5 | 6.2 | 456 | 1 | UHPT_CHLMU | Q9pj18 chlamydia m |
| 37 | 160 | 6.1 | 572 | 1 | YB30_YEAST | P38125 saccharomyc |
| 38 | 159.5 | 6.1 | 456 | 1 | UHPT_CHLTR | O84548 chlamydia t |
| 39 | 158 | 6.0 | 405 | 1 | YJHB_ECOLI | P39352 escherichia |
| 40 | 158 | 6.0 | 451 | 1 | YKJL_BACSU | P37514 bacillus su |
| 41 | 155 | 5.9 | 448 | 1 | PCAK_PSEPU | Q51955 pseudomonas |
| 42 | 154.5 | 5.9 | 543 | 1 | DAL5_YEAST | P15365 saccharomyc |
| 43 | 151.5 | 5.8 | 388 | 1 | NORA_STAAM | P21191 staphylococ |
| 44 | 149.5 | 5.7 | 396 | 1 | SOTB_ECO57 | P58529 escherichia |
| 45 | 149.5 | 5.7 | 396 | 1 | SOTB_ECOLI | P31122 escherichia |

ALIGNMENTS

```
FT TRANSMEM 144 164 POTENTIAL.
FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 205 225 POTENTIAL.
FT TRANSMEM 272 292 POTENTIAL.
FT TRANSMEM 311 331 POTENTIAL.
FT TRANSMEM 348 368 POTENTIAL.
FT TRANSMEM 375 395 POTENTIAL.
FT TRANSMEM 406 426 POTENTIAL.
FT TRANSMEM 441 461 POTENTIAL.
SQ SEQUENCE 493 AA; 52737 MW; 60792715D32553DD CRC64;

Query Match 37.0%; Score 968.5; DB 1; Length 493;
Best Local Similarity 40.5%; Pred. No. 2.9e-62;
Matches 193; Conservative 89; Mismatches 179; Indels 15; Gaps 6;

QY 27 PRAEPAPVCCSARYNLAFLSFFGFFVLYSLRVNLSVALVDMVDSNTAKDNRTSYECABH 86
DB 8 PRLVP-----STRFALSVMFFGCLVYMMRTNMSPAVVCMVNMENKTDGTGEKVSRCGE 62

QY 87 SAPIKVLHNOTGKRYWDAETOGWILGSFFGYIITQIPGGYVASRSGKLLGLGIFAT 146
DB 63 MTPVESNSSVIG-EFDWDKQTGMVLSFFYIGISQIIGHLSRYGGRVVFVTLGS 121

QY 147 AIFLFTPLAADFGVALVALALEGLGEGVTPYPAHAMWSSWAPPLERSKLISYAGA 206
DB 122 ALLTLNPAARTSEYALAILRAAIGFLOGATFPAMHTMWSVMGPPLELSVLGTGYAGA 181

QY 207 QLGTVVSPLSGVICYIY---MNTYVYFFGIVGIITWILICLVSDTPTETHKTIPTPEK 263
DB 182 QIGNVTVLPLSLGFLCEYFDGGMPSIFYIIGVGLWTAVMVYVSSDKPETHPRITPEEK 241

QY 264 EYTLSSLKWOLSSQ-----KSPWPIMPKSLPLWAIVVAHFSYNMTPTYTLTLPTVMKEV 319
DB 242 QYIVTAVEASMGKDTGKVPSTPWIKILTSFPAVWACHAGHFGAGWGYTWMVLSPLSKDV 301

QY 320 LRFNIQENGLFSAVPYLGWLCMLISGQAADNLRARWNFTLWVRVFLSLIGMIGPAIFL 379
DB 302 LGLNLSLGAVASIPYIAVFLAINAGVGLADTLRSKGILSTLNTTRAAMVALIGQIGFL 361

QY 380 VAAGFTGCDYS-LAVAFLLISTLGGFCSSGSINHLDTAPSAGILLGITNFTATIPQM 438
DB 362 VASGCGCGQDVLVIFITCGMAISGLQYAGFVWNYLEIAPPFSGTVMGTGNTISALAGI 421

QY 439 IGPILARSITPENTIGEWTFVFCIAAINVFGAIFTLFAKEGVONWA-ISPDOGH 493
DB 422 ISPAVSSYITPNTGTEQWQWLVLTAGILTGALLSFASGEVQPAKLTAECH 477
```

```
RESULT 2
Y006 CAEEL STANDARD; PRT; 576 AA.
AC P34644;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical 52.2 kDa protein ZK512.6 in chromosome III.
GN ZK512.6
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fullon L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
```

```
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -|- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z22177; CAA80150.1; -.
DR PIR; S40767; S40767.
DR WormPep; ZK512.6; CE01109.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
KW Hypothetical protein; Transmembrane; Glycoprotein; Symport;
KW Sodium transport.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 122 142 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 345 365 POTENTIAL.
FT TRANSMEM 382 402 POTENTIAL.
FT TRANSMEM 407 427 POTENTIAL.
FT TRANSMEM 438 458 POTENTIAL.
FT TRANSMEM 473 493 POTENTIAL.
SQ SEQUENCE 576 AA; 63108 MW; 71F3A9EFBE5F84CC CRC64;

Query Match 31.3%; Score 819; DB 1; Length 576;
Best Local Similarity 36.3%; Pred. No. 1.7e-51;
Matches 165; Conservative 88; Mismatches 178; Indels 24; Gaps 6;

QY 36 CSARYNLALPFSFFGFFVLYSLRVNLSVALVDMVDSNTAKDNRTSYECAHSAPIKVLHN 95
DB 66 CKRMLLAILANGFMISFGIRCNFGAKTHMYKNYTD-----PYGKVHM 110

QY 96 QTGKRYWDAETQGWILGSFFGYIITQIPGGYVASRSGKLLGLGIFATIFLTPL 155
DB 111 H---EFNWTIDELSVNMESSYFYGLVLTQIPAGFLAAKFPNKLFGGIVGAFNLILPY 167

QY 156 AADFGVGLVA-LRALEGEGVTPYPAHAMWSSWAPPLERSKLISYAGAGLGTIVSL 214
DB 168 GPKVRSYDLVAFITQIGLVQGVCPYPAHGVWRYWAPPMPERSKLATTAFTGSYAGAVLGL 227

QY 215 PLSGVICYMMWTVYVYFPGIVGIITWILICLVSDTPTETHKTIPTPEKEYI---LSSLK 271
DB 228 PLASLVSVSWAAPFYLYGCVGIWAILWFCVTEKFAFHFTISQEEKIFIEDIAGHVS 287

QY 272 NOLSSQKSPWTPMLKSLPLWAIVVAHFSYNMTFTTLTLPTVMKEVLRFRNIQENGLS 331
DB 288 NTHPTIRSPKAIYVTSKPVMAIIVANFARSWTFYLLQLNQLTYKKEALGMKIADSGLLA 347

QY 332 AVPYLGCWLCMLISGQAADNLRARWNFTLWVRVFLSLIGMIGPAIFLVAAGFICDYSL 391
DB 348 AIPHLVGCWVLMGQGLADYLRNKLSTAVRKIFNCGGFGGEAFLMIVAYTTSD-TT 406

QY 392 AVAFITISTLGGFCSSGSINHLDTAPSAGILLGITNFTATIPMIGPIIARSITPEN 451
DB 407 AIMALIAVGMSSGFAISGFNVNHLDIAPRYAALLMGFSNGICITLAGLTCFPFTEAFTAH 466

QY 452 TIGEWQTVFCIAAAINVFGAIFTLFAKEGVONWA 486
```



```
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 304 324 POTENTIAL.
FT TRANSMEM 337 356 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 429 449 POTENTIAL.
FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 465 AA; 51589 MW; C67E25A2C291EEF CRC64;

Query Match 30.7%; Score 802.5; DB 1; Length 465;
Best Local Similarity 36.5%; Pred. No. 2.1e-50;
Matches 166; Conservative 89; Mismatches 191; Indels 9; Gaps 2;

QY 33 PVCSARYNLAFSPFGFFVLYSLRNLVALVDMVDSNTAKDNRTSYECABHSAPIKV 92
Db 11 PGFCSPRYGLAILLHFCNINIVMAQRVCLNLTVMV-----NNTGSPHLSNESVVM 62
QY 93 LHNQTKKYRWDATOGWILGSFFGYIITQIPGGYVASRSRGKLLLGFGIFATAFTLF 152
Db 63 LDNKNPVSWSLDIOGLVLSVFLGMVVOAPVGLSGIYPMKRIIGSLFUSLSMLL 122
QY 153 TPLAADFGVALVALALEGLGEGVTYPAMHAMWSSWAPPLERSKLLISYAGAQLGTVV 212
Db 123 IPPAAQVGAALVIVCRVLOGIAGVSTGOHEIWKVWAPPLERGLRTSMTSGFVMPFI 182
QY 213 SLPLSGVICYMNWTVYFFFGVIGIIFWILWVRRVSLIGMIGPAIFLVAAGFIGDYSL 391
Db 183 VLVSGFICDLGWPVVFYIFGVGVLSPVFLFDDPKDHPYMSSEKDYIISLMQ 242
QY 273 QLSO-KSVPMPLKSLPLWAIWVAHFSYNTFTYLLTLLPTYMKEVLRNFQENGLS 331
Db 243 QASGQSLPIKAMKSLPLWAILNSFAFINSNLLVTPFTFSTVLHVNVRENGLS 302
QY 332 AVPLGWCWLCMILSGQAADNLRARNFSLWVRVRSGLIGMIGPAIFLVAAGFIGDYSL 391
Db 303 SLPYLLAYICGILAGQMSDFELTKRIFSVTVRKLTLLGSCFCVPVIFMCLLYLSNYS 362
QY 392 AVAFLITISTTLGGFCSSGFSINHLDIAPSYAGILLITNTFATIPGMIGPIARSUTPEN 451
Db 363 TVIFLTLANSTLSFSCGQLINALDIAPRYGFLKAVTALIGMFGGLISSTLAGLINQD 422
QY 452 TIGEMOTVFCIAAANVFGAIFFTFLFAKGEVQNW 486
Db 423 PEYAWHKISFLMAGINVTCLVFLFAKGEIQDWA 457

RESULT 5
NPTL_RAT
ID NPTL_RAT STANDARD; PRT; 465 AA.
AC Q62795;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Renal sodium-dependent phosphate transport protein 1 (Sodium/phosphate
DE cotransporter 1) (Na+/Pi cotransporter 1) (Renal sodium-phosphate
DE transport protein 1) (Renal Na(+)-dependent phosphate cotransporter
DE 1).
DE GN SLC17A1 OR NPT1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97021433; PubMed=8867793;
RA Li H., Xie Z.;
RT "Molecular cloning of two rat Na+/Pi cotransporters: evidence for
RT differential tissue expression of transcripts.";
RL Cell. Mol. Biol. Res. 41:451-460(1995).
```

```
CC CC -!- FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY.
CC CC MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA
CC CC NA+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
CC CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC CC -----
CC CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; U28504; AAC52487.1; -
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00883; sugar_cr; 1.
KW Transport; Transmembrane; Glycoprotein; Symport; Sodium transport.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 176 196 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 299 319 POTENTIAL.
FT TRANSMEM 337 357 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 431 451 POTENTIAL.
FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 465 AA; 51350 MW; 28BBBF8DC5C0AC52 CRC64;

Query Match 30.5%; Score 797.5; DB 1; Length 465;
Best Local Similarity 35.8%; Pred. No. 4.7e-50;
Matches 163; Conservative 94; Mismatches 189; Indels 9; Gaps 2;

QY 33 PVCSARYNLAFSPFGFFVLYSLRNLVALVDMVDSNTAKDNRTSYECABHSAPIKV 92
Db 11 PGFCSPRYGLAILLHFCNINIVMAQRVCLNLTVMV-----NKEPPLSNKSAEM 62
QY 93 LHNQTKKYRWDATOGWILGSFFGYIITQIPGGYVASRSRGKLLLGFGIFATAFTLF 152
Db 63 LDNKNPVSWSLDIOGLVLSVFLGMVVOAPVGLSGIYPMKRIIGSLFUSLSMLL 122
QY 153 TPLAADFGVALVALALEGLGEGVTYPAMHAMWSSWAPPLERSKLLISYAGAQLGTVV 212
Db 123 IPPAAQVGAALVIVCRVLOGIAGVSTGOHEIWKVWAPPLERGLRTSMTSGFVMPFI 182
QY 213 SLPLSGVICYMNWTVYFFFGVIGIIFWILWVRRVSLIGMIGPAIFLVAAGFIGDYSL 391
Db 183 ALLVSGFICDLGWPVVFYIFGVGVLSPVFLFDDPKDHPYMSSEKDYIISLMQ 242
QY 273 QL-SQKSVPMPLKSLPLWAIWVAHFSYNTFTYLLTLLPTYMKEVLRNFQENGLS 331
Db 243 QVHSGRQSLPIKAMKSLPLWAILNSFAFINSNLLVTPFTFSTVLHVNVRENGLS 302
QY 332 AVPLGWCWLCMILSGQAADNLRARNFSLWVRVRSGLIGMIGPAIFLVAAGFIGDYSL 391
Db 303 SLPYLLAYICGIVAGQMSDFLSRKIFSVVAVRKLTLLGFCVPVIFVVCCLLYLSNYS 362
QY 392 AVAFLITISTTLGGFCSSGFSINHLDIAPSYAGILLITNTFATIPGMIGPIARSUTPEN 451
Db 363 TVIFLTLANSTLSFSCGQLINALDIAPRYGFLKAVTALIGFGLISSTLAGLINQD 422
QY 452 TIGEMOTVFCIAAANVFGAIFFTFLFAKGEVQNW 486
Db 423 PEYAWHKNFELMAGINVTCLVFLFAKGEIQDWA 457

RESULT 6
NPTL_HUMAN
ID NPTL_HUMAN STANDARD; PRT; 465 AA.
```


Q14916; Q13783;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Renal sodium-dependent phosphate transporter protein 1 (Sodium/phosphate cotransporter 1) (Na(+)/PI cotransporter 1) (Renal sodium-phosphate transport protein 1) (Renal Na(+)-dependent phosphate cotransporter 1) (NA/PI-4).
GN SLC17A1 OR NPT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94117004; PubMed=8288239;
RA Chong S.S., Kristjansson K., Zoghbi H.Y., Hughes M.R.;
RT "Molecular cloning of the cDNA encoding a human renal sodium phosphate transport protein and its assignment to chromosome 6p21.3-p23.";
RL Genomics 18:355-359(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney cortex;
RX MEDLINE=95126933; PubMed=7826357;
RA Miyamoto K.-I., Tatsumi S., Sonoda T., Yamamoto H., Minami H., Taketani Y., Takeda E.;
RT "Cloning and functional expression of a Na(+)-dependent phosphate cotransporter from human kidney: cDNA cloning and functional expression";
RL Biochem. J. 305:81-85(1995).
RN [3]
RP FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY.
CC MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA NA+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN KIDNEY CORTEX, LIVER AND BRAIN BUT NOT IN OTHER TISSUES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X71355; CAA50490.1; ALT_INIT.
DR EMBL; D28532; BAA05888.1; -;
DR MIM; 182308; -;
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR Transport; Transmembrane; Glycoprotein; Symport; Sodium transport.
FT TRANSMEM 79 99
FT TRANSMEM 117 137
FT TRANSMEM 176 196
FT TRANSMEM 198 218
FT TRANSMEM 255 275
FT TRANSMEM 299 319
FT TRANSMEM 337 357
FT TRANSMEM 363 383
FT TRANSMEM 399 419
FT TRANSMEM 431 451
FT TRANSMEM 47 47
FT CARBOHYD 39 39
FT CARBOHYD 47 47
FT CARBOHYD 56 56
FT CONFLICT 35 36
FT CONFLICT 77 77
FT CONFLICT 207 207
FT CONFLICT 231 231
SQ SEQUENCE 465 AA; 50884 MW; 687A19F245C4ED16 CRC64;

Query Match 30.2%; Score 790.5; DB 1; Length 465;
Best Local Similarity 37.7%; Pred. No. 1.5e-49;
Matches 177; Conservative 78; Mismatches 177; Indels 37; Gaps 6;
QY 33 PVCCARYNLAFLSFFGVLYSLRVNLSVALVDVDS-----NTTAKDNRTSYEAE 85
DB 11 PGFCFRYGLSLVHCCNVIIITAQRACNLNLTWVWVNVSTDPHGLPNTSTK----- 60
QY 86 HSAPIKVLHNTQKKYRWDATQGWILGSFFGYITITQIPGCVASRSGKLLGLGIFA 145
DB 61 -----KLLDNKNPMYNSPDQIILSTSGVILIIQVPYGFSGIYTKKMGFALCL 115
QY 146 TAIFTLFTPLAADFGVALVALRLEGEGVTYPAMHAMSSWAPPLERSKLLSTISYAG 205
DB 116 SSVLSLLIPPAAGIGVAVWVVCRAVQGAAGIVATAQAEIYVVKVAPPLERGLTSMSTG 175
QY 206 AQLGTVVSLPLSGVICYNNWTYVFFGIVGIIIFILWICLVSDTPETHKTITPYEREY 265
DB 176 FLGSPFIVLLVTGVICSLGNPMVFIIFGACCAVCLLWFLVFDYDDPKDHPGISSEKEY 235
QY 266 ILSLKNOL-SSQKVPWIPMLKSLPLWAIWVAFHSYNWTFYTLTLTLPYMKVELRNI 324
DB 236 ITSSLVQQVSSRSQSLPIKAILKSLPVWALSIGSTFFWSHNMITYTPMFINSMLHVN 295
QY 325 QENGFLSAVPLGCLWCLMILSGQADNLNRWNSFTLWVRVSLIGMIGPAIFLVAAGF 384
DB 296 KENGLFSLPYLFAWICGNLQGLSDFELTRNLSLVIAVRKLTAAAGFLPAIFGVCLPY 355
QY 385 ICDYSLAVAFITSTTGGFCSSGFSINHLDIASVAGILGINTTATIGMIGPIIA 444
DB 356 LSTFYSIVIFILAGATGSCFGLGVFINGLDIAPRYGFI----KACSTLGMIGGLIA 411
QY 445 RSLT-----PENTIGEWOTVFCIAAINVFGAIFFTLFAKEGVQNW 486
DB 412 SLTGLILKQDPESA---WFKTFILMAINVTGLIFYLIVATAEIQDWA 457
RESULT 7
YRT3_CAEEL STANDARD; PRT; 544 AA.
AC Q10046;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 60.5 kDa protein T07A5.3 in chromosome III.
GN T07A5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Buck D.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -!- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z48055; CAA88134.1; -;
DR WormPep; T07A5.3; CE01648.
KW Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport;
KW Sodium transport.
FT TRANSMEM 50 70
FT TRANSMEM 105 125
FT TRANSMEM 128 148

```

FT TRANSMEM 155 175 POTENTIAL.
FT TRANSMEM 194 214 POTENTIAL.
FT TRANSMEM 221 241 POTENTIAL.
FT TRANSMEM 286 306 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 361 381 POTENTIAL.
FT TRANSMEM 385 405 POTENTIAL.
FT TRANSMEM 417 437 POTENTIAL.
FT TRANSMEM 451 471 POTENTIAL.
SQ SEQUENCE 544 AA; 60543 MW; 4615B3C27D9F86CC CRC64;

Query Match 29.3%; Score 766.5; DB 1; Length 544;
Best Local Similarity 35.6%; Pred. No. 9.2e-48;
Matches 164; Conservative 90; Mismatches 166; Indels 41; Gaps 9;

QY 39 RYNLAFLSFFGVLYSLRVNLVALVDMVDSNTTAKDRTSYECAHSAPIKVLINQGT 98
DB 47 RWOIATLAHFGFAISFGRNFGVAKNRNVNFTDAYGE-----VHE--- 88
QY 99 KKYRDAETOGTGLSGFFGYIITQIPGGYVASRSGKLLGFGIFATIFTLETPLAAD 158
DB 89 REFLWTGAEVGMWESFFGYRAAQIPAGVLAAKFAPNKFIMLGILVASFMIILSAISFN 148
QY 159 FGVA---LVALRALEGLGEGVTYPAMHAMWSSWAPPLERSKILSISYAGAQLGTGVVSLP 215
DB 149 FHPYTDIFVMVQVQVGLALGVLYPAMHGVWKFAPPLERSKLATTAFTGSSGVVMTGLP 208
QY 216 LSGVICYNNWIVYVFFGVIGIILFIMICLVSDPETHKTIPTPYEKEYILS-----SL 270
DB 209 ASAYLSHFSWSPTFFVFGVGLIINSLIMWYVSSHSPETHGYISDDKKQVTEKIGDVAV 268
QY 271 KNOLSSQKSVPTPMLKSLPLWAIVVAHFSYNNVFTLLTLLPTMYKEVLRNIOENGLF 330
DB 269 KNM--SLTILPHRDMTSSAVWAIITCTCRSGWGFLLNGNLQTYMKDVLHDIKNSGFI 326
QY 331 SAVPYLGCWMLCMLSQAAADNLRARNWFTLWVRRVFSLIGMIGPAIFLVAAGFIGC--- 387
DB 327 SIPPQGMCIIVTLATQGLCDYLRSCKMSTEAVRKSVNTFG-----FTVEAMMLGCLAF 380
QY 388 --DYSLAVAFLTISTTLGFCSSGFSINHLDTAPSAGILLGITNTFATIPGMIGPIIAR 445
DB 381 VRDPVIAVCLVITACVSGSVLSGFGVNVNHFDAIPRYAPILMTGANGLAGVAG--VGMVVTN 439
QY 446 SLTPENTIGEMOTVFCIAAINVFGAIFFTFLAKGVQWNA 486
DB 440 TTYIQNPDG--WKWVFLAMADIFGVIFFLIFAKGDVLPWA 479

RESULT 8
NPT3 HUMAN
ID NPT3 HUMAN STANDARD; PRT; 436 AA.
AC O00624;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Sodium-dependent phosphate transport protein 3 (Sodium/phosphate
DE cotransporter 3) (Na(+)/PI cotransporter 3).
GN SLC17A3 OR NPT3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed-9149941;
RA Ruddy D.A., Kronmal G.S., Lee V.K., Mintier G.A., Quintana L.,
RA Domingo R. Jr., Meyer N.C., Irrinki A., McClelland E.B., Fullan A.,
RA Mapa F.A., Moore T., Thomas W., Loeb D.B., Harmon C., Tsuchihashi Z.,
RA Wolff R.K., Schatzman R.C., Feder J.N.;
RT "A 1.1-Mb transcript map of the hereditary hemochromatosis locus.";
RL Genome Res. 7:441-456(1997).
CC -!- FUNCTION: MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE

```

```

CC INTO CELLS VIA NA+ COTRANSPORT (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U91328; AAB82085.1; -
DR EMBL; U90544; AAB53422.1; -
KW Transport; Transmembrane; Glycoprotein; Phosphorylation; Symport;
KW Sodium transport.
FT TRANSMEM 26 46 POTENTIAL.
FT TRANSMEM 98 118 POTENTIAL.
FT TRANSMEM 130 150 POTENTIAL.
FT TRANSMEM 183 203 POTENTIAL.
FT TRANSMEM 211 231 POTENTIAL.
FT TRANSMEM 273 293 POTENTIAL.
FT TRANSMEM 317 337 POTENTIAL.
FT TRANSMEM 366 406 POTENTIAL.
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 69 69 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 436 AA; 47350 MW; DF02F618E83A572F CRC64;

Query Match 25.7%; Score 672; DB 1; Length 436;
Best Local Similarity 38.9%; Pred. No. 4.2e-41;
Matches 139; Conservative 69; Mismatches 133; Indels 16; Gaps 4;

QY 27 PRAEPAPVCCSARYNLAFSLPFGFFVLYSLRVNLVALVDMVDSNTTAKDRTSYECAH 86
DB 5 PATRKGPDPCSLRYGLALIMHFSNFTMTITQRVSLTALIMV--NTQOQGLSN---AST 59
QY 87 SAPIKVLHNOT-----GKKYRWDAETOGWILGSFFGYIITQIPGGYVASRSGK 136
DB 60 EGPVADAFNNSSISKEFDTKASVQWSPETOGIIFSSINYGILLIIFSGVLGIFGAK 119
QY 137 LLLGFGIFATIFTLTPLAADFGVAGVALRALEGLGEGVTYPAMHAMWSSWAPPLERS 196
DB 120 KMLGAGLLISLLTLTFTPLAADFGVILVIMVTVQGMAGMAWTGQFTIWAkWAPPLERS 179
QY 197 KLLSISYAGAQLGTGVVSLPGLSGVICYMMWTVVFFFGVIGIILFIMICLVSDPETHK 256
DB 180 KLTITAGSGSAFGSFILCVGGLISQALSQWPFIFYIFGSGTCVCCLLWFTVIYDDPMHP 239
QY 257 TITPYEKEYILSLKNQLSSQ--KSPVWIPMLKSLPLWAIVVAHFSYNNVFTYLLTLPY 315
DB 240 CLSVREKEHILSLAQOQSPSPGRAVPIKAWVTCLPLWAIFLPGFFSHFWLCTIITLPIY 299
QY 316 MKEVLRFNQENGFLSAVPLGCLWCLMSGQAADNLRARNWFTLWVRRVFSLIGM 372
DB 300 ISTLLHVNIRDSGVLSLSPFIAAASCTILGGQLADFLLSRNLRLITVRKLLSSLOM 356

RESULT 9
NPT4 HUMAN
ID NPT4 HUMAN STANDARD; PRT; 401 AA.
AC O00476;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Sodium-dependent phosphate transport protein 4 (Sodium/phosphate
DE cotransporter 4) (Na(+)/PI cotransporter 4).
GN SLC17A4 OR NPT4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```


CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: BELONGS TO THE PHTHALATE PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D26185; BAA05188.1; -;
DR EMBL; Z99124; CAB16094.1; -;
DR Subtilist; BG10016; ybbo.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr_1;
KW Hypothetical protein; Transmembrane; Transport; Complete proteome.
FT TRANSMEM 26 46 POTENTIAL.
FT TRANSMEM 61 81 POTENTIAL.
FT TRANSMEM 96 116 POTENTIAL.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 150 170 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 242 262 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
FT TRANSMEM 347 367 POTENTIAL.
FT TRANSMEM 385 405 POTENTIAL.
FT TRANSMEM 407 427 POTENTIAL.
SQ SEQUENCE 435 AA; 48248 MW; 9630914D9B606208 CRC64;

Query Match 12.0%; Score 314; DB 1; Length 435;
Best Local Similarity 25.3%; Pred No. 1.8e-15;
Matches 116; Conservative 71; Mismatches 192; Indels 80; Gaps 18;
Qy 49 GFFVL-YSLRVNLSVALVDMVDSNTAKDNRYSYCAEHSAPIKVLHNQTKGKYRWDAAET 107
Db 29 GIILNFDRAISVAIPAQDS-----FHLTATE 58
Qy 108 QGWLIGSEFFGYITQPGGVASRSGKLLGFGIFATAITFTPLAADFGVALVAL 167
Db 59 LGIVFSIYTSYTLMLPVGSLDRFGVAVTRVGMTIWSFLTIL--LAFLLGKLLLYLF 116
Qy 168 RALEGLSGGVTPAMHAMWSSWAPLERSKLLSISYAGAOQLGTVVSLPSGVICYNNWT 227
Db 117 RFLGLTSASAPPAASKATALWFPSPERGLANSFLDSAAKFSNVIGAPLVAFVLTTFDWR 176
Qy 228 YVYFFGIVGIWILWICLVSDTPETHKTIPTVEKEYILSKLNQLSSQKSVPMI--PM 285
Db 177 VAFLTICINVL-FTIFWQYEQPERHKRISKSELNVI---QKHNAITTEQIPYKTGPL 232
Qy 286 LKSL-----PLMAIVAHFSYNWTFYTLTLPTTYMKEVLRNIQENGFLSAVYPYLGWLC 341
Db 233 LKLFTRKVKWGLMIGETGYTFTNLLTLPTFKHTYGMGLMSSGLTAVP---WLI 288
Qy 342 MILSQADNLRAW-----NFTLWVRVFLSMIGMGAIFLVAAFGICDYS LAV 393
Db 289 STISGIAVGG-----WLVDYFIKKGYPNKTYRTVIIVGMSGFFFL---GSI-LTNNTIV 340
Qy 394 AFLTISTTLGGFCSS---GFSINHLDIAP----SYAGILGLITNTFTATIPGMIGPIARS 446
Db 341 AIIICISIGLAGISATAPVGSIS-AELAPIGSVSMLSMWNLAN-----NLFSGIITAS 393
Qy 447 LTPE--NTIGEMQVFCIAAIVNFGAIFFTFLFAKGEVQ 483
Db 394 LTGYLFDVTSFTLSFLVAGFVLLGLVFEY-VFVLGDVK 431

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 15, 2002, 08:25:02 ; Search time 53.59 Seconds
(without alignments)
1597.919 Million cell updates/sec

Title: US-09-776-865-4

Perfect score: 2617
Sequence: 1 MKSPVSDLPDSGEGSDRT.....LFAKGEVQNAISDHQHRN 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-----------|--------------------|
| 1 | 2617 | 100.0 | 495 | 6 Q9WZD1 | Q9mzd1 ovis aries |
| 2 | 2329 | 89.0 | 495 | 4 Q9UGH0 | Q9ugh0 homo sapien |
| 3 | 2329 | 89.0 | 536 | 4 Q9NRA2 | Q9nra2 homo sapien |
| 4 | 991.5 | 37.9 | 559 | 5 Q9VYG7 | Q9vyg7 drosophila |
| 5 | 985.5 | 37.7 | 502 | 5 Q9VDM0 | Q9vdm0 drosophila |
| 6 | 965 | 36.9 | 582 | 11 Q9J1I2 | Q9j1i2 rattus norv |
| 7 | 965 | 36.9 | 582 | 11 Q920B7 | Q920b7 mus musculu |
| 8 | 962.5 | 36.8 | 582 | 4 Q9P2U8 | Q9p2u8 homo sapien |
| 9 | 941 | 36.0 | 529 | 5 Q9V7S5 | Q9v7s5 drosophila |
| 10 | 936 | 35.8 | 586 | 5 Q23514 | Q23514 caenorhabdi |
| 11 | 932 | 35.6 | 560 | 4 Q9P2U7 | Q9p2u7 homo sapien |
| 12 | 927 | 35.4 | 560 | 11 Q62634 | Q62634 rattus norv |
| 13 | 926 | 35.4 | 483 | 5 Q61369 | Q61369 drosophila |
| 14 | 884 | 33.8 | 466 | 4 Q96LH1 | Q96lh1 homo sapien |
| 15 | 874.5 | 33.4 | 497 | 4 Q9Y2C5 | Q9y2c5 homo sapien |
| 16 | 859.5 | 32.8 | 481 | 5 Q9VJW8 | Q9vjw8 drosophila |

| | | | | | |
|----|-------|------|-----|-----------|--------------------|
| 17 | 847.5 | 32.4 | 453 | 5 Q9NKF8 | Q9nkf8 drosophila |
| 18 | 845 | 32.3 | 479 | 5 Q9W4G7 | Q9w4g7 drosophila |
| 19 | 835.5 | 31.9 | 496 | 5 Q9VKC9 | Q9vkc9 drosophila |
| 20 | 819 | 31.3 | 563 | 5 Q9TZN7 | Q9tzn7 caenorhabdi |
| 21 | 807.5 | 30.9 | 465 | 11 Q91WV5 | Q91wv5 mus musculu |
| 22 | 798 | 30.5 | 465 | 5 Q9V905 | Q9v905 drosophila |
| 23 | 792.5 | 30.3 | 573 | 5 Q09932 | Q09932 caenorhabdi |
| 24 | 788.5 | 30.1 | 467 | 4 Q9H531 | Q9h531 homo sapien |
| 25 | 787 | 30.1 | 512 | 5 Q9VFX2 | Q9vfx2 drosophila |
| 26 | 784.5 | 30.0 | 497 | 5 Q9V8M8 | Q9v8m8 drosophila |
| 27 | 767.5 | 29.3 | 462 | 5 Q9V8M9 | Q9v8m9 drosophila |
| 28 | 760 | 29.0 | 493 | 5 Q9VR44 | Q9vr44 drosophila |
| 29 | 744 | 28.4 | 512 | 10 Q82390 | Q82390 arabidopsis |
| 30 | 737 | 28.2 | 475 | 5 Q9V8N0 | Q9v8n0 drosophila |
| 31 | 737 | 28.2 | 491 | 5 Q9V8N1 | Q9v8n1 drosophila |
| 32 | 716 | 27.4 | 529 | 10 Q9SDI4 | Q9sdi4 oryza sativ |
| 33 | 715.5 | 27.3 | 560 | 5 Q9VQC0 | Q9vcq0 drosophila |
| 34 | 697 | 26.6 | 449 | 5 Q9V334 | Q9v334 drosophila |
| 35 | 696.5 | 26.6 | 390 | 11 Q921B2 | Q921b2 mus musculu |
| 36 | 696.5 | 26.6 | 524 | 5 Q94886 | Q94886 drosophila |
| 37 | 696.5 | 26.6 | 524 | 5 Q9V763 | Q9v763 drosophila |
| 38 | 618 | 23.6 | 495 | 5 Q9W1Z0 | Q9w1z0 drosophila |
| 39 | 606.5 | 23.2 | 420 | 4 Q9H533 | Q9h533 homo sapien |
| 40 | 585 | 22.4 | 527 | 5 Q16923 | Q16923 caenorhabdi |
| 41 | 580.5 | 22.2 | 444 | 5 Q9W1Z1 | Q9w1z1 drosophila |
| 42 | 577 | 22.0 | 468 | 5 Q9V9J0 | Q9v9j0 drosophila |
| 43 | 571.5 | 21.8 | 432 | 10 Q9FKV1 | Q9fkvl arabidopsis |
| 44 | 569 | 21.7 | 792 | 5 Q9W1Z3 | Q9w1z3 drosophila |
| 45 | 561.5 | 21.5 | 413 | 10 Q23065 | Q23065 arabidopsis |

ALIGNMENTS

```
RESULT 1
Q9WZD1 ID Q9MZD1 PRELIMINARY; PRT; 495 AA.
AC Q9MZD1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE MEMBRANE GLYCOPROTEIN SP55.
GN SP55.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Fu C., Bardhan S., Cetateanu N.D., Lloyd S.R., Yan H., Carter C.E.,
RA Shi E., Venkov C., Yakes M.F., Page D.L., H.C.G.;
RT "Identification of a novel membrane protein from mammalian cells that
RT interact with the anti-pathoangioic compound CM101."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF244578; AAF9770.1; -.
SQ SEQUENCE 495 AA; 54536 MW; 649D7C4A59B28272 CRC64;
```

Query Match 100.0%; Score 2617; DB 6; Length 495;
Best Local Similarity 100.0%; Pred. No. 2.1e-197;
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MKSPVSDLPDSGEGSDRTPLQRPAPRAPAPVCCSARYNLAFSLFFGFVLYSLRVNL 60
Db 1 MKSPVSDLPDSGEGSDRTPLQRPAPRAPAPVCCSARYNLAFSLFFGFVLYSLRVNL 60
QY 61 SVALVDMVDSNTAKNRTSYECAHSAPTKVLHNTGKKYRWDAETOGWILGSFFGYI 120
Db 61 SVALVDMVDSNTAKNRTSYECAHSAPTKVLHNTGKKYRWDAETOGWILGSFFGYI 120
QY 121 ITQIPGGYVASRSGKLLLGFGIFATAIFTLTPLAADFVGVALVALLEGLEGVTP 180
```

```
Db 121 ITQIPGGYVASRSGKLLGFGIFATAIFTLFTPLAADFGVGALVALRALEGEGVTYP 180
|||||
QY 181 AMHAMSSWAPPLERSKLLSISYAGAQGLGVVWSLPSGVICYMNTYVYFFGIVGIW 240
|||||
Db 181 AMHAMSSWAPPLERSKLLSISYAGAQGLGVVWSLPSGVICYMNTYVYFFGIVGIW 240
|||||
QY 241 FILWICLVSDTPETHKITPYEKEYILSSLNKQLSSQKSPWPIMLKSLPLMAIVVAHFS 300
|||||
Db 241 FILWICLVSDTPETHKITPYEKEYILSSLNKQLSSQKSPWPIMLKSLPLMAIVVAHFS 300
|||||
QY 301 YNWTFTLLTLLPTYMKEVLRNIQENGFLSAVPYLGWCMLCMTLSGQAADNLRARWNFT 360
|||||
Db 301 YNWTFTLLTLLPTYMKEVLRNIQENGFLSAVPYLGWCMLCMTLSGQAADNLRARWNFT 360
|||||
QY 361 LWVRRVFLSLGIMGPAIFLVAAGFICDYSLAVAFLTISTTLGGFCSSGFSINHLDIAPS 420
|||||
Db 361 LWVRRVFLSLGIMGPAIFLVAAGFICDYSLAVAFLTISTTLGGFCSSGFSINHLDIAPS 420
|||||
QY 421 YAGILLGINTNTFATIPGMIPIIARSILTPENTIGEQTVFCIAAAINVFCAIFFTLFAKG 480
|||||
Db 421 YAGILLGINTNTFATIPGMIPIIARSILTPENTIGEQTVFCIAAAINVFCAIFFTLFAKG 480
|||||
QY 481 EVONWAISSDHQHRN 495
|||||
Db 481 EVONWAISSDHQHRN 495
|||||

RESULT 2
Q9UGH0 PRELIMINARY; PRT; 495 AA.
ID Q9UGH0
AC Q9UGH0
DT 01-OCT-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE MEMBRANE GLYCOPROTEIN HP59.
DE SIALIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20047778; PubMed=10581036;
RA Verheijen F.W., Verbeek E., Aula N., Beerens C.E.M.T., Havelaar M.C.,
RA Joosse M., Peltonen L., Aula P., Galjaard H., Van der Spek P.J.V.D.,
RA Mancini G.M.S.;
RT "A new gene, encoding an anion transporter, is mutated in sialic acid
RT storage diseases.";
RL Nat. Genet. 23:462-465(1999).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL; AJ387747; CAB62540.1; -
DR InterPro; IPR003662; sub.transpporter.
DR Pfam; PF00083; sugar_tr; 1.
KW Transmembrane.
SQ SEQUENCE 495 AA; 54639 MW; 5C6C154B3E93A19E CRC64;
```

Query Match 89.0%; Score 2329; DB 4; Length 495;
Best Local Similarity 86.3%; Pred. No. 8.5e-175;
Matches 427; Conservative 39; Mismatches 29; Indels 0; Gaps 0;

```
QY 1 MKSPVSDLAPSDGEGSDRTPLQAPRAEPAPVCCSARYNLAFSLFFGFFVLYSLRVNL 60
|-||| ||| :||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MRSPVRLDARNDEESTDRTPLPGAPRAEAPVCCSARYNLALAFFGFFIVYALRVNL 60
|||||
QY 61 SVALVDMVDSNTTAKDNRTSYECAHSAPIKVLHNQTKGKRWDAETQGWILGSFFGYI 120
|||||
Db 61 SVALVDMVDSNTTLEDNRTSKACPEHSAPIKVHNNQTKGKQWDAETQGWILGSFFGYI 120
|||||
QY 121 ITQIPGGYVASRSGKLLGFGIFATAIFTLFTPLAADFGVGALVALRALEGEGVTYP 180
|||||
Db 121 ITQIPGGYVASKIGKMLLGFGILGTAFTLTPTIAADLGVGPLIVLRALEGEGVTYP 180
|||||
```

```
QY 181 AMHAMSSWAPPLERSKLLSISYAGAQGLGVVWSLPSGVICYMNTYVYFFGIVGIW 240
|||||
Db 181 AMHAMSSWAPPLERSKLLSISYAGAQGLGVVWSLPSGVICYMNTYVYFFGIVGIW 240
|||||
QY 241 FILWICLVSDTPETHKITPYEKEYILSSLNKQLSSQKSPWPIMLKSLPLMAIVVAHFS 300
|||||
Db 241 FILWICLVSDTPETHKITPYEKEYILSSLNKQLSSQKSPWPIMLKSLPLMAIVVAHFS 300
|||||
QY 301 YNWTFTLLTLLPTYMKEVLRNIQENGFLSAVPYLGWCMLCMTLSGQAADNLRARWNFT 360
|||||
Db 301 YNWTFTLLTLLPTYMKEVLRNIQENGFLSAVPYLGWCMLCMTLSGQAADNLRARWNFT 360
|||||
QY 361 LWVRRVFLSLGIMGPAIFLVAAGFICDYSLAVAFLTISTTLGGFCSSGFSINHLDIAPS 420
|||||
Db 361 LWVRRVFLSLGIMGPAIFLVAAGFICDYSLAVAFLTISTTLGGFCSSGFSINHLDIAPS 420
|||||
QY 421 YAGILLGINTNTFATIPGMIPIIARSILTPENTIGEQTVFCIAAAINVFCAIFFTLFAKG 480
|||||
Db 421 YAGILLGINTNTFATIPGMIPIIARSILTPENTIGEQTVFCIAAAINVFCAIFFTLFAKG 480
|||||
QY 481 EVONWAISSDHQHRN 495
|||||
Db 481 EVONWAISSDHQHRN 495
|||||

RESULT 3
Q9NRA2 PRELIMINARY; PRT; 536 AA.
ID Q9NRA2
AC Q9NRA2
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE MEMBRANE GLYCOPROTEIN HP59.
DE HP59.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Fu C., Bardhan S., Cetateanu N.D., Lloyd S.R., Yan H., Carter C.E.,
RA Shi E., Venkov C., Yakes M.F., Page D.L., Hellerqvist C.G.;
RT "Identification of a novel membrane protein from mammalian cells that
RT interacts with the anti-pathogenic compound CM101.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF244577; AAF97769.1; -.
SQ SEQUENCE 536 AA; 58721 MW; 3DDADFCC84007724 CRC64;
```

Query Match 89.0%; Score 2329; DB 4; Length 536;
Best Local Similarity 86.3%; Pred. No. 9.3e-175;
Matches 427; Conservative 39; Mismatches 29; Indels 0; Gaps 0;

```
QY 1 MKSPVSDLAPSDGEGSDRTPLQAPRAEPAPVCCSARYNLAFSLFFGFFVLYSLRVNL 60
|-||| ||| :||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 42 MRSPVRLDARNDEESTDRTPLPGAPRAEAPVCCSARYNLALAFFGFFIVYALRVNL 101
|||||
QY 61 SVALVDMVDSNTTAKDNRTSYECAHSAPIKVLHNQTKGKRWDAETQGWILGSFFGYI 120
|||||
Db 102 SVALVDMVDSNTTLEDNRTSKACPEHSAPIKVHNNQTKGKQWDAETQGWILGSFFGYI 161
|||||
QY 121 ITQIPGGYVASRSGKLLGFGIFATAIFTLFTPLAADFGVGALVALRALEGEGVTYP 180
|||||
Db 162 ITQIPGGYVASKIGKMLLGFGILGTAFTLTPTIAADLGVGPLIVLRALEGEGVTYP 221
|||||
QY 181 AMHAMSSWAPPLERSKLLSISYAGAQGLGVVWSLPSGVICYMNTYVYFFGIVGIW 240
|||||
Db 222 AMHAMSSWAPPLERSKLLSISYAGAQGLGVVWSLPSGVICYMNTYVYFFGIVGIW 281
|||||
QY 241 FILWICLVSDTPETHKITPYEKEYILSSLNKQLSSQKSPWPIMLKSLPLMAIVVAHFS 300
|||||
Db 282 FILWICLVSDTPETHKITPYEKEYILSSLNKQLSSQKSPWPIMLKSLPLMAIVVAHFS 341
|||||
```

[illegible]

| | | | | |
|----|-----|------------|------|--|
| Qy | 476 | LFAKGEVQNW | 486. | |
| | | | | |
| Db | 496 | LFASGEKQPW | 506 | |

| | | |
|------------------|---|---------------------------|
| RESULT | 8 | |
| Q9P2U8 | | |
| ID | Q9P2U8 | PRELIMINARY; PRT; 582 AA. |
| AC | Q9P2U8; | |
| DT | 01-OCT-2000 (TREMBlrel. 15, Created) | |
| DT | 01-OCT-2000 (TREMBlrel. 15, Last sequence update) | |
| DT | 01-OCT-2000 (TREMBlrel. 15, Last annotation update) | |
| DE | DIFFERENTIATION-ASSOCIATED NA-DEPENDENT INORGANIC PHOSPHATE | |
| DE | COTRANSPORTER. | |
| GN | DNPI. | |
| OS | Homo sapiens (Human). | |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | |
| NCBI_TaxID=9606; | | |
| RN | [1] | |
| RP | SEQUENCE FROM N.A. | |
| RC | TISSUE-BRAIN THALAMUS; | |
| RA | Atahara Y., Mashima H., Onda H., Hisano S., Kasuya H., Hori T., | |
| RA | Yamada S., Tomura H., Yamada Y., Inoue I., Kojima I., Takeda J.; | |
| RT | "Molecular cloning of a novel brain-type Na ⁺ -dependent inorganic | |
| RT | phosphate cotransporter." | |
| RL | J. Neurochem. 0:0-0(2000). | |
| DR | EMBL; AB032435; BAA92874.1; -. | |
| SQ | SEQUENCE 582 AA; G64392 MW; CE761E56FA18C6AD CRG64; | |

```
Query Match      36.8%; Score 962.5; DB 4; Length 582;
Best Local Similarity 41.4%; Pred. No. 2e-67;
Matches 193; Conservative 85; Mismatches 161; Indels 27; Gaps 6;
```

| | | | |
|----|-----|---|-----|
| Qy | 32 | APVC-CSA-----RYNLAFLSFFGFVLYSURVLNSVALVDMDVNNTAKDNRTSYECAE | 85 |
| | | : : : : : : : : : : : : | |
| Db | 57 | APLDCDTCFGLPRYYITAIMSGLGFCISFGIRCNLGAIVDMWNSTIHRGGKVIKEKA- | 115 |
| | | : : : : : : : : : : : : | |
| Qy | 86 | HSAPIKVLUHNOTGKKYRWDATQTGWILGSFPFYGIITQPGYVASRSSGKLILGFGIFA | 145 |
| | | : : : : : : : : : : : : | |
| Db | 116 | -----KFNWDPETVMINGHSFWFGYIIITQPGYTASRLAANRVFAGAILL | 161 |
| | | : : : : : : : : : : : : | |

| | | | |
|--------|-------------|--|--------------|
| Db | 116 | -----KFNDPETHGHSFFWGYIITQIPGGYIASRLAANRVFGAAILL | 161 |
| Qy | 146 | TAIFTLFTPLAADFGVALRALRLEGIGGVYTPAMHAMWSSWAPPLERSKLLTSYAG | 205 |
| Db | 162 | TSTLNMLIPSAARVHYGCVFVRILQGLVGVYTPACHGTSWKPPLERSRLATTSFCG | 221 |
| Qy | 206 | AOLGTVVSLPUSGVICYMMWTVYVFFGVIGIWFILMCLVSDTPETHKTTIPYEKEY | 265 |
| Db | 222 | SYAGAVIAMPLAGILVQYTGWSSVYVYGVGMMVFMWLLVSYESPAKHPITTDERRY | 281 |
| Qy | 266 | ILSSL---KNOLSSQ--SVYPWIPMLKSLPLWALIVAHFYSYNTFYTLTLPTMYKEVL | 320 |
| Db | 282 | TEESIGBSANLLGAMKFKTPWRKFFTSMPYIAIVANFCRSWTFYLLILISQPAYFEFV | 341 |
| Qy | 321 | RFNTQENGFLSAPYLGWCWLCMILSGOADMRLARWNFSTLWVRVPSFLTCMGIPAFV | 380 |
| Db | 342 | GFETSKVGLMSAVPHLVMTIIVPGIGQIADFLRSKQILSTTTVRKIMNCGGFGMEATLL | 401 |
| Qy | 381 | AAGFIGDCDYSLAVAFLLTISTTLGGFCSSGSFINHLDIAPSYAGILLGITNTFATIPMG | 440 |
| Db | 402 | VVGY-SHTRGVAISFLVLAVGFSGFALSGFNVNHLDIAPRYASILMGSINGVGVTLSGVC | 460 |
| Qy | 441 | PIIARSLTPENTIGEWQTVCIAAAINVFALFTFLAKGEVQWA | 486 |
| Db | 461 | PIIVGAMTKNSREEWQYVFLIAALVHYGVIFYAIFASGEKQPWA | 506 |
| RESULT | 9 | | |
| ID | Q9V7S5 | PRELIMINARY; | PRT; 529 AA. |
| AC | Q9V7S5; | Q9V7S6; | O61364; |
| DT | 01-MAY-2000 | (Tremblrel. 13, Created) | |


```
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBSJ databases.
KW EMBL; U58737; AAK39396.1; -.
KW Hypothetical protein.
SQ SEQUENCE 586 AA; 65045 MW; 95D52B7F7FA7DD0F CRC64;

Query Match 35.8%; Score 936; DB 5; Length 586;
Best Local Similarity 42.0%; Pred. No. 2.4e-65;
Matches 192; Conservative 82; Mismatches 155; Indels 28; Gaps 8;

QY 39 RYNLAFLSFFGVLYSLRVNLVALVDMVDSNTAKDNKTSYECASAPIKVLNQTG 98
Db 145 RHVVAILALGFANFIYAMRANLSIAIVEMT-SGTERKVNCT-----LH-VLG 190
QY 99 KKYRWDATQGTLSGFFGYIITQIPGGVVASRSRGKLLGFGIFATAFTLTPLAAD 158
Db 191 DFNWTPMTQGVVLSFFGYIVSQLPGGYLATHGAKTIFAGTGTAVFTLTTPPEAR 250
QY 159 FGVALVALRALGEGVTPYPAHAMWSSWAPPLERSKILLSYAGAOIGTVVSLPLSG 218
Db 251 MGYGMLVFARFMGLEGVTPYPAHWISRWAPPMQTKLATFAFSSYFGTVVAMPLSA 310
QY 219 VICYYMNTYVFFGVIGLWILMICLVSDTPETHKITPYEKEYIILSSLNQLSQK 278
Db 311 YLGEHFGWPMIEFFGALGVIMCMWYKTVHDPEDDPDKISTSE----LALLQDAVSQ 366
QY 279 S--VPMTPLKSLPLVAIVVAHESYNWTFVTLTLTPYMKVLRNIQENGELSAPYL 336
Db 367 HYIVPQAQLRSKPWAWIVAHSAQNLGFYIMLTNLPKMKLDIAGYNVEKAGIASLPY 426
QY 337 GCWLCMILSQQAADNLRARWNFSTLWVRVFLSGLIGMIGPAIFLVAAAGFGCDYS 393
Db 427 LMGFOITGGQLCDYLRDKHYDTLFRKMACALGFIGQSVEL----FLVMTSLSLLV 482
QY 394 AFUTISTTLGGFCSSGFSINHLDIAPSYAGILGINTNTATIPGMIGPIIARSLTP 533
Db 483 LFPSISIGLGGICWCFSVNLHDLAPQYAGHLMATSNATIPGIFGLLVGAIVQNG 542
QY 454 GEMQTVFCIAAANVFGAIFFTLFAKGEVQNWAISDH 490
Db 543 GEMNVIMYIIISAYILGAIAFWKPADATLQPAWA-AEH 578

RESULT 11
Q9P207 PRELIMINARY; PRT; 560 AA.
AC Q9P207
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE BRAIN-SPECIFIC NA-DEPENDENT INORGANIC PHOSPHATE COTRANSPORTER.
GN ENPI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Aihara Y., Mashima H., Onda H., Hisano S., Kasuya H., Hori T.,
RA Yamada S., Tomura H., Yamada Y., Inoue I., Kojima I., Takeda J.;
RT "Molecular cloning of a novel brain-type Na+-dependent inorganic
RT phosphate cotransporter.";
RL J. Neurochem. 0:0-0(2000).
DR EMBL; AB032436; BAA92875.1; -.
SQ SEQUENCE 560 AA; 61613 MW; C88DAFB34B6E45B6 CRC64;

Query Match 35.6%; Score 932; DB 4; Length 560;
Best Local Similarity 39.5%; Pred. No. 4.7e-65;
Matches 191; Conservative 92; Mismatches 171; Indels 30; Gaps 6;

RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBSJ databases.
KW EMBL; U58737; AAK39396.1; -.
KW Hypothetical protein.
SQ SEQUENCE 586 AA; 65045 MW; 95D52B7F7FA7DD0F CRC64;

Query Match 35.8%; Score 936; DB 5; Length 586;
Best Local Similarity 42.0%; Pred. No. 2.4e-65;
Matches 192; Conservative 82; Mismatches 155; Indels 28; Gaps 8;

QY 39 RYNLAFLSFFGVLYSLRVNLVALVDMVDSNTAKDNKTSYECASAPIKVLNQTG 98
Db 145 RHVVAILALGFANFIYAMRANLSIAIVEMT-SGTERKVNCT-----LH-VLG 190
QY 99 KKYRWDATQGTLSGFFGYIITQIPGGVVASRSRGKLLGFGIFATAFTLTPLAAD 158
Db 191 DFNWTPMTQGVVLSFFGYIVSQLPGGYLATHGAKTIFAGTGTAVFTLTTPPEAR 250
QY 159 FGVALVALRALGEGVTPYPAHAMWSSWAPPLERSKILLSYAGAOIGTVVSLPLSG 218
Db 251 MGYGMLVFARFMGLEGVTPYPAHWISRWAPPMQTKLATFAFSSYFGTVVAMPLSA 310
QY 219 VICYYMNTYVFFGVIGLWILMICLVSDTPETHKITPYEKEYIILSSLNQLSQK 278
Db 311 YLGEHFGWPMIEFFGALGVIMCMWYKTVHDPEDDPDKISTSE----LALLQDAVSQ 366
QY 279 S--VPMTPLKSLPLVAIVVAHESYNWTFVTLTLTPYMKVLRNIQENGELSAPYL 336
Db 367 HYIVPQAQLRSKPWAWIVAHSAQNLGFYIMLTNLPKMKLDIAGYNVEKAGIASLPY 426
QY 337 GCWLCMILSQQAADNLRARWNFSTLWVRVFLSGLIGMIGPAIFLVAAAGFGCDYS 393
Db 427 LMGFOITGGQLCDYLRDKHYDTLFRKMACALGFIGQSVEL----FLVMTSLSLLV 482
QY 394 AFUTISTTLGGFCSSGFSINHLDIAPSYAGILGINTNTATIPGMIGPIIARSLTP 533
Db 483 LFPSISIGLGGICWCFSVNLHDLAPQYAGHLMATSNATIPGIFGLLVGAIVQNG 542
QY 454 GEMQTVFCIAAANVFGAIFFTLFAKGEVQNWAISDH 490
Db 543 GEMNVIMYIIISAYILGAIAFWKPADATLQPAWA-AEH 578

RESULT 11
Q9P207 PRELIMINARY; PRT; 560 AA.
AC Q9P207
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE BRAIN-SPECIFIC NA-DEPENDENT INORGANIC PHOSPHATE COTRANSPORTER.
GN ENPI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Aihara Y., Mashima H., Onda H., Hisano S., Kasuya H., Hori T.,
RA Yamada S., Tomura H., Yamada Y., Inoue I., Kojima I., Takeda J.;
RT "Molecular cloning of a novel brain-type Na+-dependent inorganic
RT phosphate cotransporter.";
RL J. Neurochem. 0:0-0(2000).
DR EMBL; AB032436; BAA92875.1; -.
SQ SEQUENCE 560 AA; 61613 MW; C88DAFB34B6E45B6 CRC64;

Query Match 35.4%; Score 927; DB 11; Length 560;
Best Local Similarity 39.4%; Pred. No. 1.2e-64;
Matches 184; Conservative 89; Mismatches 168; Indels 26; Gaps 4;

QY 30 EPAPVCCSA-----RYNLAFLSFFGVLYSLRVNLVALVDMVDSNTAKDNKTSYEC 84
Db 48 DPPVVDCTCGLPRIIATMSGLGFCISFGIRCNLGAIVSMVNSTTHRGHVVQKA 107
```

```
QY 85 EHSAPIKVLHNQTKKRYWDAETOGWILGFFYGYIITQIPGGYVASRSGGKLLGLGFIF 144
Db 108 Q-----FNWDPEIVGLHGFFWGYIVTQIPGGFICQKFAANRVFGFAIV 152
QY 145 ATAITFTPLAADFGVALVALRALGLEGGVYTPAMHAMWSSWAPPLERSKLLSISYA 204
Db 153 ATSTLNLMLPSAARVHYGCVIFRVLQGLVEGVTYPACHGIWKPAPPLERSRLATTAF 212
QY 205 GAOLCTVWSPLSGVICYNNWNTYVFFYFGLIIFWILWICLVSDTPETHKTIITYEKE 264
Db 213 GSYAGAVVAMPLAGVLQVSGNSVYVYVGSFGIFWYLFLLVSYESPALHPSISEERK 272
QY 265 YILSSKLNQSSOKS-----PHPMKLSPLWAIIVVAHFSYNWNTYFTLLTLLPTMYKEV 319
Db 273 YIEDAGESAKLNPYKFTNPRRFTSPVYAIIVANFCRSWTYLLISOPATFEV 332
QY 320 LRFNIQENGFLSAPVYLGWCLMILSQQAADNLARWNFTLWVRVFSLIGMIGPAIFL 379
Db 333 FGEISKVGLVSALPHLWMTIIVPIGQIADFLRSRHINSTTNVRKLMNCGFGFMEATLL 392
QY 380 VAGFTGCDYSLAVAFITSTTLGGCCSGFSINHLDIAPSYAGILLGTTNTFATPGMI 439
Db 393 LVVGY-SHSGVAISFLVLAVGSFGAISGFNHNHLDIAPRYASILMGISNGVGTLSGMV 451
QY 440 GPIIARSLTPENTIGEMQTVFCIAAAINVFCAIFFTLFAKGEVQNW 486
Db 452 CPIIVGANTHKTRWEQVYFLASLVHYGVYFYGVFASGEKQPNW 498
RESULT 13
O61369 PRELIMINARY; PRT; 483 AA.
AC O61369
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE INORGANIC PHOSPHATE COTRANSPORTER.
GN PICOT.
OS Drosophila ananassae (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7217;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TAI 13-1610;
RA Da Lage J.-L., Alland C.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -/- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
DR EMBL: AF024691; AAC39088.1; -.
DR FlyBase: FBgn0024472; Dana\Picot.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
KW Transport; Transmembrane.
FT TRANSMEM 187 207 POTENTIAL.
FT TRANSMEM 292 312 POTENTIAL.
FT TRANSMEM 383 403 POTENTIAL.
FT TRANSMEM 420 440 POTENTIAL.
SQ SEQUENCE 483 AA; 52885 MW; 6832750611B775F0 CRC64;

Query Match 35.4%; Score 926; DB 5; Length 483;
Best Local Similarity 42.3%; Pred. No. 1.2e-64;
Matches 190; Conservative 81; Mismatches 164; Indels 14; Gaps 8;

QY 47 FFGFFVLSLRVLSVALVDMVDSNTAKDNRTSY--ECAHSAPIKVLHNQTKKRYWMD 104
Db 3 FLGMANAYVMTNMSVAIVAMV-NHTAISKGEEDDECGDRDIPID--DSQDG-EFPWN 58
QY 105 AETQGHILGSFFGYIITQIPGGYVASRSGGKLLGLGFIFATFAITFLTPLA--DFGVGA 163
Db 59 AALQGYILSFFGYIITQIPGILAKKYGSLRFLGYGLINSVFAFLVPAAREGVNG 118
```

```
QY 164 LVALRALEGLGEGVTPAMHAMWSSWAPPLERSKLLSISVAGAOCTVYVSLPSLVICVY 223
Db 119 LCARVRIQGLGEGVTPVCTHAMLAWKWIPNERSRGAAYVAGAQFGTIISMPLSGLLAEY 178
QY 224 ---MNNVYVYFVGIIWIFILWICLVSDTPETHKTIITYEKEYILSSL-KNLSSQKS 279
Db 179 GFDGWPISFYVFGIVGTVMSTIAFLIFVEDPSTHPKIDEREKKYNESLWGTDVIKSP 238
QY 280 VPWIPMLKSLPLWAIIVVAHFSYNWNTYFTLLTLLPTMYKEVLRNIOENGFLSAPVYLGW 339
Db 239 IPFKSVTKSLPFAILFAHMGHNYGYETLMTLPTMYKQVLRFSKNSGLSLSLPYLAMW 298
QY 340 LCMILSQQAADNLARWNFTLWVRVFSLIGMIGPAIFLVAAGFTGCDYSLAVAFITIS 399
Db 299 LLSMFTSVTADMISKRFSLSFATRIINSIGQYGLALIAASYGCDRALTLAILITG 358
QY 400 TTLGGCCSGFSINHLDIAPSYAGILLGTTNTFATPGMIGPIIARSLTPENT---IGEW 456
Db 359 VGLNGGIYSGFKINHLDLTPRFAGELMSITNCSANLAGLLAPIAAGNLISDPSKPYMGQW 418
QY 457 QTVFCTAAAINVFCAIFFTLFAKGEVQNW 485
Db 419 QIVFFTAFAFYIICGTGYFNYFGSGERQFW 447
RESULT 14
Q96LH1 PRELIMINARY; PRT; 466 AA.
AC Q96LH1
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE DJ139G21.1.2 (SODIUM PHOSPHATE SOLUTE CARRIER FAMILY 17 MEMBER 2
DE (ISOFORM 2)) (FRAGMENT).
GN SLC17A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Peck A.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL138726; CAC69102.1; -.
FT NON_TER 1
SQ SEQUENCE 466 AA; 50664 MW; F64CE5FC89E5FC3D CRC64;
```

Query Match 33.8%; Score 884; DB 4; Length 466;
Best Local Similarity 38.7%; Pred. No. 2.2e-61;
Matches 179; Conservative 91; Mismatches 176; Indels 16; Gaps 4;

```
QY 36 CSARYNLAFSPFGFVYLSLRVLSVALVDMVDSNTAKDNRTSYECAHSAPIKVLHN 95
Db 2 CSLRYGLALIMHFSNFTMITQVSLSIATIAMV--NTQOQGLSN---ASTEGPVADAEN 56
QY 96 QT-----GKKYRWDAETQGWILGSSFFGYIITQIPGGYVASRSGKLLGLGFIFA 145
Db 57 NSSISIKFEFTKASYQWSPETQGIIFSSINYGIIITLIPSGYLAFGAKMGLAGLLI 116
QY 146 TAIFTLTPLAADFGVALVALRALGLEGGVYTPAMHAMWSSWAPPLERSKLLSISYAG 205
Db 117 SLLTFTPLAADFGVILVIMVTVTQVMAQGMATWQFTIWAQWAPPLERSKLTITAGSG 176
QY 206 AOLGTVVSPLSGVICYNNWNTYVFFYFGLIIFWILWICLVSDTPETHKTIITYEKEY 265
Db 177 SAFSGFIILCVGLISQALSQALSWPFIYIFGSTGCVCCLLWFTVYDDPMHHPICISVREKH 236
QY 266 ILSLKNQSSQ-KSVYPWIPMLKSLPLWAIIVVAHFSYNWNTYFTLLTLLPTMYKEVLRNI 324
Db 237 ILSLQAQSPSPGRAVPIKAMVTCPLWAIIFLGFHFWLCTIILTYLPTYISTLLHVN 296
```


QY 325 QENGFLSAVPIYGCWLCMLSCQAADNLPRNFTLWVRVFSILGMIGPAIFLVAAGF 384
Db 297 RDSGVLSSLPFTAAASCTILGQADLLSRNLLRLITVRKLFSSGLGLLPSCAVALPF 356
QY 385 IGCXYSLAVAFITSTLGGFCSSGFSINHLDIAPSAGILLGITNTFATIPGMIGPIIA 444
Db 357 VASSVYIITILLIPGYSNLCDSGFIINTLDIAPRYASFLMGSRGFLIAGIISSTAT 416
QY 445 RSLTPENTIGEQTVFCIAAAINVFGAIFFTLFAKGEVQNA 486
Db 417 GFLISQDFSGWRNFFLSAAVNMFLGYFLFTFGQAEIQDWA 458

RESULT 15
QY9Y2C5 PRELIMINARY; PRT; 497 AA.
AC Q9Y2C5;
DT 01-NOV-1999 (TremBLrel. 12, Created)
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE NA/PO4 COTRANSPORTER HOMOLOG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99253143; PubMed=10319585;
RA Shibui A., Tsunoda T., Seki N., Suzuki Y., Sugane K., Sugano S.;
RT "Isolation and chromosomal mapping of a novel human gene showing
RT homology to Na+/PO4 cotransporter.";
RL J. Hum. Genet. 44:190-192(1999).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL: AB020527; BAA76663.1; -
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
KW Transmembrane.
SQ SEQUENCE 497 AA; 54055 MW; 5A5616AA52D4990D CRC64;

Query Match 33.4%; Score 874.5; DB 4; Length 497;
Best Local Similarity 37.4%; Pred. No. 1.3e-60;
Matches 185; Conservative 101; Mismatches 188; Indels 21; Gaps 7;

QY 1 MKSPYDLAPSDGEGSDRTPLLQAPRAEPAPVCCSARYNLAEISFGFFVLYSLRVNL 60
Db 7 VKATVGDIS-SDGN-----LNVAQECRKFGFCVSRHGLALILQLCNFSIITQQMNL 57

QY 61 SVALVDMVDSNTAKDNRTSYECASHSAPIKVLHNQTKK-----YRWDAETQGWILG 113
Db 58 SIAPAMV--NNTAPPSPNASTERPSTDGQYWNETLKEFKAMAPAYDWSPEIQGIILS 115

QY 114 SFPGYIITQIPGGYVSRSGKLLLGIFATAIFTFTPLAADFGVALVALRALEGL 173
Db 116 SLNYGSFLAPIPSGVAGIFGAKYVVGAGLFSSELTFLFIPLAANAGVALLIVLRIVQGI 175

QY 174 GEGVTYPAMHAMWSWAPLERSKLLISYAGAOLGVVYVPLSGVICVYMNMTVYVFF 233
Db 176 AQVMVLTQYSIHWKAPLERSQITTAGSGMLGSFIVLLAGLLCQTICGWYVYIF 235

QY 234 GIVGIWFIILICLVSDPTPEHTITPYEKEYILSKNQ-LSSQKSVPIWPKLSPLW 292
Db 236 GGIGCACCPFLWPLYIDDPVNHPTISAGEKRYIVCSLAQQDCSPGWSLPIRAMIKSLPLW 295

QY 293 AIVVAHFSYNWTFYLLTLLPTMYKEVLRFNQENGLSANVPY-LGCWLCMLLSQAADN 351
Db 296 ALLVSFYCEYMLFTYIMATPTTYSIVLQANLRDSGILSALPFVYGC-ICILGLGLADF 354

QY 352 LRARNFSTLWVRVFSILGMIGPAIFLVAAGFIGCDYSLAVAFLTISTTLGGFCSSGFS 411
Db 355 LLRSKILRLIIRKLTAGLVFPSPVILVLPWVRSSHMTWTFVLVLSAITSFCESGAL 414

QY 412 INHLDIAPSYAGILLGITNTFATIPGMIGPIIARSLTPTENTIGEQTVFCIAAAINVFGA 471
Db 415 VNFLDIAPRYTGLKGLLQVFAHIAAGISPTAAGFFISQDSEFGWRNVFLLSAAVNISGL 474
QY 472 IFFTLLFAKGEVQNA 486
Db 475 VFYLIFGRADVQDWA 489

Search completed: July 15, 2002, 08:25:03
Job time: 168 sec

| Sequence | Strd | Orig | ZScore | EScore | Len | Documentation |
|-------------------|------|---------|---------|----------|--------|-----------------------------------|
| gb_om:AF244578 | + | 2617.00 | 4124.27 | 2.3e-221 | 2844 | ! A224578 Ovis aries membrane g |
| gb_om:AX0207626 | + | 2617.00 | 4124.27 | 2.3e-221 | 2844 | ! AX0207626 Sequence 3 from Paten |
| gb_pat:AX138494 | + | 2329.00 | 3668.89 | 5.4e-196 | 2512 | ! AX138494 Sequence 2 from Paten |
| gb_pat:HSAB387747 | + | 2329.00 | 3668.89 | 5.4e-196 | 2512 | ! AX387747 Homo sapiens mRNA for |
| gb_pat:AX0207624 | + | 2329.00 | 3667.31 | 6.6e-196 | 2930 | ! AX0207624 Sequence 1 from Paten |
| gb_pat:AF244577 | + | 2329.00 | 3667.31 | 6.6e-196 | 2930 | ! AF244577 Homo sapiens membran |
| gb_pr:BC020961 | + | 2329.00 | 3665.99 | 7.8e-196 | 3329 | ! BC020961 Homo sapiens, solute |
| gb_pr:BC060776 | + | 985.50 | 1341.83 | 1.6e-77 | 1841 | ! BC060776 Drosophila melanogast |
| gb_htg:AC014246 | + | 981.50 | 1505.63 | 1.7e-75 | 33448 | ! AC014246 Drosophila melanogast |
| gb_in:AC023685 | + | 981.50 | 1489.01 | 1.4e-74 | 167928 | ! AC023685 Drosophila melanoga |
| gb_in:AC023711 | + | 981.50 | 1488.64 | 1.5e-74 | 171463 | ! AC023711 Drosophila melanoga |
| gb_in:AE003491 | + | 981.50 | 1482.26 | 3.4e-74 | 323461 | ! AE003491 Drosophila melanoga |
| gb_ro:AF324864 | + | 969.00 | 1512.41 | 7.0e-75 | 2528 | ! AF324864 Mus musculus vesiculic |
| gb_ro:AF271235 | + | 968.00 | 1506.14 | 1.6e-75 | 3982 | ! AF271235 Rattus norvegicus dif |
| gb_pr:AB032435 | + | 963.50 | 1499.10 | 3.9e-75 | 3946 | ! AB032435 Homo sapiens DNPI mRN |
| gb_in:AY069501 | + | 946.50 | 1475.87 | 7.6e-74 | 2748 | ! AY069501 Drosophila melanogast |
| gb_ro:RNU07609 | + | 933.50 | 1458.41 | 7.2e-73 | 2024 | ! U07609 Rattus norvegicus brain |
| gb_pr:AB032436 | + | 932.00 | 1454.42 | 1.2e-72 | 2366 | ! AB032436 Homo sapiens BNPI mRN |
| gb_pat:IA0028 | + | 931.00 | 1451.41 | 1.8e-72 | 2716 | ! IA0028 Sequence 1 from patent |
| gb_pat:IA0029 | + | 931.00 | 1451.41 | 1.8e-72 | 2716 | ! IA0029 Sequence 1 from patent |
| gb_pat:IA0220 | + | 931.00 | 1451.41 | 1.8e-72 | 2716 | ! IA0220 Sequence 3 from patent |
| gb_pat:IA0221 | + | 931.00 | 1451.41 | 1.8e-72 | 2716 | ! IA0221 Sequence 3 from patent |
| gb_pat:IA73259 | + | 931.00 | 1451.41 | 1.8e-72 | 2716 | ! IA73259 Sequence 1 from patent |
| gb_pat:IA73260 | + | 931.00 | 1451.41 | 1.8e-72 | 2716 | ! IA73260 Sequence 3 from patent |
| gb_in:AF024691 | + | 907.50 | 1401.04 | 1.1e-69 | 9703 | ! AF024691 Drosophila ananassae |
| gb_htg:AC020077 | + | 903.00 | 1377.50 | 2.3e-68 | 47723 | ! AC020077 Drosophila melanogast |
| gb_in:AC091501 | + | 903.00 | 1362.83 | 2.3e-67 | 198390 | ! AC091501 Drosophila melanoga |
| gb_in:AC003806 | + | 903.00 | 1359.72 | 3.2e-67 | 268299 | ! AC003806 Drosophila melanoga |
| gb_in:CEC38C10 | + | 892.00 | 1363.49 | 1.4e-67 | 34193 | ! C19153 Caenorhabditis elegans |
| gb_ro:BC018306 | + | 885.50 | 1383.09 | 1.1e-68 | 1874 | ! BC018306 Mus musculus, Similar |
| gb_pr:AB020527 | + | 874.50 | 1362.18 | 1.6e-67 | 2626 | ! AB020527 Homo sapiens mRNA for |
| gb_pr:AB024903 | + | 864.50 | 1343.02 | 1.9e-66 | 3616 | ! AB024903 Homo sapiens CDNA: FI |
| gb_in:AY075277 | + | 859.50 | 1342.81 | 2.0e-66 | 1710 | ! AY075277 Drosophila melanogast |
| gb_htg:AC014600 | + | 859.50 | 1311.95 | 1.0e-64 | 34229 | ! AC014600 Drosophila melanogast |
| gb_in:AE003640 | + | 859.50 | 1290.77 | 1.6e-63 | 267488 | ! AE003640 Drosophila melanoga |
| gb_in:DROSADH01 | + | 859.50 | 1289.46 | 1.8e-63 | 303885 | ! AE003407 Drosophila melanoga |
| gb_in:AC023715 | + | 845.50 | 1273.83 | 1.4e-62 | 160622 | ! AC023715 Drosophila melanoga |
| gb_in:AE003432 | + | 845.50 | 1266.74 | 3.4e-62 | 319551 | ! AE003432 Drosophila melanoga |
| gb_htg:AC013869 | + | 829.50 | 1275.55 | 1.1e-62 | 11568 | ! AC013869 Drosophila melanoga |
| gb_htg:AC020252 | + | 827.00 | 1259.66 | 8.4e-62 | 36842 | ! AC020252 Drosophila melanoga |

VGALVALRALEGLGEGVTYPAMHAMSSWAPLERSKLLSISYAGAQLGTVVSLPLSG
VICYVNMNTYVFYFGIIVGIWFIWILMCLIVSDTPETHKTIPTPEXEYILSSKLNOLSS
OKSVPIWMLKSLPLWAIWIAHFSYNWTFYLLTLPTMYKVELVRNIOENGFLSAVP
YLGCVKADNLISGOADNLRARNFSTLWRRVFSILGIMGPAIFLVAAGFIGCDYSLA
VAFILISLTLGFCSSGFSINHLIDAPSAGILLGINTFNATIPGMIGFPIIARSLTPE
NTIGSMQVFEICAAAINVFGAIFFFLFAKGEVQNWAIASHOQHRN"

BASE COUNT 745 a 623 c 594 g 882 t
ORIGIN

alignment_scores:
Quality: 2617.00 Length: 495
Ratio: 5.287 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-776-865-4 x AX207626 ..

Align seg 1/1 to: AX207626 from: 1 to: 2844

1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGluGlySe 17
|||||
84 ATGAAGTCCCGGTTTCGGACTAGCCCCGAGCGACGGCGAGGAGGCTC 133
17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProV 34
|||||
134 GGACCGCACACCGCTCTCGACGCGCGCGCGGGGGAACCCGCTCCAG 183
34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyPhe 50
|||||
184 TATGCTGCTCTGCTCGTGAACACCTAGCATTTTGTCTCTTTTGGTTTC 233
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
|||||
234 TTCGTTCTCTATTCATTACCGGTGAATCTGAGCGTTGCACCTAGTGGACAT 283
67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
|||||
284 GGTGGATTCAACACAACTGCCAAAGATAATAGAACCTCCTACGAGTGTG 333
84 laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
|||||
334 CAGAGCATCTGCTCCCATAAAAGTTCTTCAACACCAACCGGGTAAAAAG 383
101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTy 117
|||||
384 TACCGGTGGGATCGAGAACTCAAGATGGATTCCTGGATCTTTTCTA 433
117 rGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSerg 134
|||||
434 TGGCTACATCATCACAAATTCCTGGAGGATATGTTGCCAGCAGAAAGTG 483
134 lyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
|||||
484 GGGGGAAGCTGTGCTAGGATTCGGGATCTTGCACAGCTATCTTCACC 533
151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
|||||
534 CTGTTCACTCCCTCGCTGCAGATTCGGAGTCGGAGCCCTTGTGGCACT 583
167 uArgAlaLeuGluGlyLeuGlyGluGlyValThrTyrProAlaMethHisA 184
|||||
584 CAGGGCACTAGAAGGGCTAGGAGGGGTGCACATATCCAGCCATGCATG 633
184 laMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
|||||
634 CCATGTGGTCTTATGGGCTCCCGCTCTTGAAGAAGCAAGCTCTCGAGT 683
201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
|||||
684 ATTTTCATATCAGAGGCACAACTTGGACAGTAGTTTCTCTCTCTCTTC 733
217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234

|||||
734 TGGAGTAATTTGCTACTATATGAATTGCACTTATGTCTCTTCTATTTCTTTG 783
234 lyIleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
|||||
784 GCATTTGTTGAATCATCTGGTTTATTTTATGGATCTGCTTAGTAGTATG 833
251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
|||||
834 ACACCAGAAACTCACAGACAATCACTCCCTATGAAAGGAGGATATATCT 883
267 uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284
|||||
884 TTCATCATTAATAAATCAGCTCTCTTACAGAACTCAGTGCCTGGATAC 933
284 roMetLeuLysSerLeuProLeuTrpAlaIleValValAlaHisPheSer 300
|||||
934 CTATGCTGAAATCACTGCCACTTTGGGCTATTGCTGTGCACATTTTCT 983
301 TyrAsnTrpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLy 317
|||||
984 TACAACCTGGACTTTTATACCTTTGTTGACCTTATGCTACTTACATGAA 1033
317 sGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValP 334
|||||
1034 GGAAGTCTCTAAGGTTCAATATTAAGAGAAATGGGTTTTTATCTGCACTC 1083
334 roTyrLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 350
|||||
1084 CTTATTTAGGTGTTGTTGTTATGATGATCTCTGGGTCAAGCTGGTGAC 1133
351 AsnLeuArgAlaArgTrpAsnPheSerThrLeuTrpValArgArgValPh 367
|||||
1134 AYTTAAGGCCAAGATGGAATTTTCAACTCTGTGGGTTTCAAGAGTTTT 1183
367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
|||||
1184 TAGCCTTATAGGGATGATTGGACCTGCGATATCTCGTTGCCGCGAGGAT 1233
384 heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
|||||
1234 TTATAGGCTGTGATTATTCCTTGGCTGTTCATTCCTTAACCATATCAACA 1283
401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIl 417
|||||
1284 ACCCTGGGAGGCTTTTCTCTCTCTGGATTTAGCATCAACCATCTGGACAT 1333
417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaT 434
|||||
1334 TGCTCCTTCGTATGCTGGTATTCCTTGGGATCACAATAACCTTTGCCA 1383
434 hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
|||||
1384 CTATTCCTGGAATGATTGGGCCCATCATTCGCAAGTCTTACCCCTGAG 1433
451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaAlaIleAs 467
|||||
1434 AACACTATTGGAGAAATGGCAAACTGTTTTCGATCGCTGCTGCTATCAA 1483
467 nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnA 484
|||||
1484 TGTATTTGGTGCCATTTTCTTCACTATTCGCAAGTGAAGTGAAGTGCAA 1533
484 snTrpAlaIleSerAspHisGlnGlyHisArgAsn 495
1534 ACTGGGCCATCAGTGATCACCAAGGACACAGAAAC 1568

seq_name: gb_pat:AX138494

seq_documentation_block:

LOCUS AX138494

DEFINITION Sequence 2 from Patent EP1069184.

ACCESSION AX138494

linear

DNA

2512 bp

PAT 30-MAY-2001

```
VERSION      AX138494.1  GI:14274389
KEYWORDS
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 2512)
AUTHORS      Human anion transporter gene implicated in salla disease and
              lysosomal sialic acid transport
TITLE        Patent: EP 1069184-A 2 17-JAN-2001;
              Akzo Nobel N.V. (NL)
JOURNAL
FEATURES
source       Location/Qualifiers
              1..2512
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
BASE COUNT   612 a 585 c 566 g 749 t
ORIGIN
alignment_scores
  Quality: 2329.00      Length: 495
  Ratio: 4.903          Gaps: 0
  Percent Similarity: 95.960  Percent Identity: 86.263
alignment_block
US-09-776-865-4 x AX138494 ..
Align seg 1/1 to: AX138494 from: 1 to: 2512
1  MethySerProValSerAspLeuAlaProSerAspGlyGluGluGlySe 17
|||||
271  ATGAGGCTCCGGTTCGAGACCTGCCCGGAACGATGGCGAGGAGCAC 320
17  rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaPro 34
|||||
321  GGACCGCAGCGCTCTTCTACCGCGGCCCGCCACGGCGCGAGCGCTCCAG 370
34  aLysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyPhe 50
|||||
371  TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
51  PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
|||||
421  TTCATTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 470
67  tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
|||||
471  GGTAGATTCAAATCAAACTTTAGAGAGATAATAGAACTTCCAAAGCGGTGC 520
84  LaGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
|||||
521  CAGAGCATTCGCTCCCAATAAAGTTTCATCATATCAACCGGTAAGAAG 570
101  TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTy 117
|||||
571  TACCAATGGGATGAGAACTCAAGAGTGGATTCTCGGTTCTCTTTTAA 620
117  rGlyTyrIleThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134
|||||
621  TGGCTACATCATCATCAGATTCCTGGAGGATATGTTGCCAGCAATAAG 670
134  lyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
|||||
671  GGGGAAATGCTGCTAGGATTTGGATCTTGGCACTGGTGTCTCACC 720
151  LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
|||||
721  CTGTTTCATCCCATTTGTCAGATTTAGAGAGTTGGACCACTCAATGTACT 770
167  uArgAlaLeuGluGlyLeuGlyGluGlyValThrTyrProAlaMethHisA 184
|||||
771  CAGAGCACTAGAGGACTAGGAGAGGGGTGTACATTTCCAGCCCATGCATG 820
```

```
184  laMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
|||||
821  CCATGTGTCTCTCTTGGCTCCCTCTTGAAGAAGCAAACTTCTTAGC 870
201  IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
|||||
871  ATTTTCATATGAGAGACAGCTTGGGACAGTAATTTCTCTCTCTCTTC 920
217  rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
|||||
921  TGGAAATAATTTGCTACTATATGAATGGACTATGTCTCTTACTTTTTC 970
234  lyIleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
|||||
971  GTACTATTTGGAAATATTTGGTTCTTTTGGATCTGGTTAGTTAGTGAC 1020
251  ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
|||||
1021  ACACCACAAAACACAAGAGAATTTCCCATTTATGAAAAGGAATACATTCT 1070
267  uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284
|||||
1071  TTCAATCATTAAGAAATCAGCTTTCTTCAAGAACTCAGTCCCGTGGGTAC 1120
284  roMetLeuLysSerLeuProLeuTrpAlaIleValAlaHisPheSer 300
|||||
1121  CCATTTTAAATCCCTGCCACTTTGGGCTATCGTAGTTGCACACTTTTCT 1170
301  TyrAsnTrpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLy 317
|||||
1171  TACAACCTGGACTTTTATATCTTTATGACATTTATGCTACTTATATGAA 1220
317  sGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValP 334
|||||
1221  GGAGATCCTTAAGGTCAATGTTCAGAGAAATGGGTTTTTATCTTCATGTC 1270
334  roTyrLeuGlyCysTrpTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 350
|||||
1271  CTATATTAGGCTCTTGGTTATGATGATCCTGCTGCTCAAGCTGCTGAC 1320
351  AsnLeuArgAlaArgTrpAsnPheSerThrLeuTrpValArgArgValPh 367
|||||
1321  AATTTAAGGGCAAAATGGAATTTTCAACTTTATGTGTGCGAGAATTTT 1370
367  eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
|||||
1371  TAGCCTTATAGGAATGATGACCTGCAGATTCCTGGTAGCTGCTGGCT 1420
384  heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
|||||
1421  TCATTGGCTGTGATTATTTTGGCCGCTGCTTTCCCTAACTATATCAACA 1470
401  ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIl 417
|||||
1471  ACACCTGGAGGCTTTTGTCTCTCTGAGATTTAGCATCAACCATCTGGAT 1520
417  eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaT 434
|||||
1521  TGTCTCTTGTATGCTGATCCTCCCTGGGCAATCACAATAATACATTGCCA 1570
434  hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
|||||
1571  CTATTCCAGGAATGGTTGGCCCGCTCATTTGCTAAAGCTCTGACCCCTGAT 1620
451  AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs 467
|||||
1621  AACACTGTTGGAGAAATGGCAACCGTGTCTATATTTGCTGCTATATAA 1670
467  nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnA 484
|||||
1671  TGTTTTTGGTGCCATTTCTTTACACTATTTCGCCAAAGGTGAAGTACAAA 1720
```


|||||:::||||||| |||||:::| |
1321 AATTTAAGGCAAAATGAATTTTTTCAACTTATTGTGTGCAGAAATTTT 1370

367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaLaGlyP 384
1371 TAGCCTTATAGGAATGATTGGACCTGCAGTATTCTCGTAGCTGCCT 1420

384 heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
1421 TCATTGGCTGTGATTATTCCTTTGGCCGPTGCTTCTTCACTATATATCAACA 1470

401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspII 417
1471 ACACITGGGAGGCTTTTGCTCTCTTGGAATTAGCATCAACCATCTGGATAT 1520

417 eaLProSerTyRAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlat 434
1521 TGCCTCTCGTATGCTGGTATCCTCTGGGCATCACAAATACATTGCCA 1570

434 hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
1571 CTAATCCAGGAATGTTGGGCCGCTCATPTGCTAAAAGTCTGACCCCTGAT 1620

451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs 467
1621 AACACTGTTGGAGAATGGCAACCCTGTTCTATTATTGCTGCTGCTATTAA 1670

467 nValPheGlyAlaIlePheThrLeuPheAlaLySGlyGluValGlnA 484
1671 TGCTTTTGGTGCCATTTCCTTACACTATTGCCCAAAGGTGAAGTACAAA 1720

484 snTrpAlaIleSerAspHisGlnGlyHisArgAsn 495
1721 ACTGGGCTCTCAATGATCACCATTGGACACAGACAC 1755

seq_name: gb_pat:AX207624

```

seq_documentation_block:
LOCUS      AX207624                               2930 bp    DNA    linear    PAT 31-AUG-2001
DEFINITION Sequence 1 from Patent WO0156598.
ACCESSION  AX207624
VERSION     AX207624.1  GI:15422329
KEYWORDS   human.
SOURCE     Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 2930)
            Hellerqvist,C.G.
            Methods for preventing or attenuating pathoangiogenic conditions by
            using the gbs-toxin (cm101) receptor as a vaccine
            Patent: WO 0156598-A 1 09-AUG-2001;
            VANDERBILT UNIVERSITY (US)
FEATURES   Location/Qualifiers
            1..2930
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                263..1873
                    /note="unnamed protein product"
                    /codon_start=1
                    /protein_id="CAC60358.1"
                    /db_xref="GI:15422330"
                    .
                    /translation="MAAGATWPPRPVQPPAPGGFGLSGRRSLCQVASTPAHVGVNVRMS
                    PVRLNDGGEESTDRTPLLPGPAEAAAPCCSARYNLAILAFFGFFVYVALRVNLN
                    VALIVMVDNSNTLEADRTSKCPESHAPIKVHNQTKGYQWDAETQGWILGSGFFYGY
                    ITIQPGVYVASK IGGKMLGIFGLTAVLTFTPTAADLGVGLPLVLRALGLEIGGVY
                    TFPAMHAWSSWAPLRLKSLISYAGAOLGTVISLPSGLIICYYNNWTVYVFFGT
                    IGFWEFLWLWVSDTPQKHRTSHVEKEYILSLRNOLSSQKSPVWPVTLKSLPLWA
                    IVVAHESYNNWTFYTLTLPTYNKEILRNVOENGFLSLPYLGSRLCMTLSQQAADN
                    LRKNWFSLCVARIESLIGMIGCPVFLVAAGFIGCDYSLAVAFLLITSTLGGFCSSG
                    FSNLHDIAPSYAGILIGLITNTFATIPGMVGPVIAKSLTPDNTVGEWQTVFYTAAAIN
                    VFGAIEFTTLFAKEGVQNWALNDHGHRR"
            729 a 698 c 681 g 822 t
BASE COUNT

```


/co

```

transporter), member 5"
/protein_id="AAH20961.1"
/db_xref="GI:18088703"
/translation="MSPVRLDIARNDEGSTDRTPLLPGAPRAEAAVCCSARYNLAI
LAFEGFVIYALRVNLISVALVDVMSNTILEDNRISKACPEHSAPIKVHHNQTKYIQ
WDDETQGWILGSFFYITITPGGVASKIGKMLLGFGLGTAVLTTLTPPIAADLG
VGLIVLIRALEGGEFFGTPPPHAMHMSWAPLERKSLISYAGAQELVLTSLPSLG
ITCYNNWVVFVEFGTIGIFWFLMWLVSDTPQPKHKRISHYEKQELIISLRNQLS
QKSPWPVPLIKSLPLWAIYVAHFSYNTVITLLTLPYMKETLRNVOENGFLSSLP
YLGSLWMLSGOAAADNLKANNFSTLCVRRIFSLGIMGICPAFLVYAAFGICGDSLA
VAFLTISTTLGFCSSGSPSINHDIAPSLAGILLGTNTFATIPMGVGPVIAKSLTPD
NTVGEWQTVFYTAAAINVFAGIIFTLFAKGEQVONALNDHHGRH"
863 a 718 c 727 q 1021 t
BASE COUNT

```

```

alignment_scores:
  Quality: 2329.00      Length: 495
  Ratio: 4.903         Gaps: 0
  Percent Similarity: 95.960  Percent Identity: 86.263

alignment_block:
  US-09-776-865-4 x BC020961

  Align seg 1/1 to: BC020961 from: 1 to: 3329

1 MetLysSerProValSerAspLeuAlaProSerAspGlyClyclucGlySe 17
  |||:::|||||
125 ATGAGGCTCCGGTTCGAGACCTGGCCCGGAGCATGGCGAGGAGAGAC 174
  |||:::|||||

17 rAspArgThrProLeuLeuGlnArgAlaLaProArgAlaGluProAlaPro 34
  |||:::|||||
175 GGACCGCAGCGCTCTTCTACGGGCGGCCACAGGGCGGAGCGCTCCAG 224
  |||:::|||||

34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPhe 50
  |||:::|||||
225 TGTGCTGCTCGTCGCTGACAACTTACGAATTTGGCTTTTGGTTTC 274
  |||:::|||||

51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 57
  |||:::|||||
275 TTCATTGCTGATGCAATTCAGTGTGAATCTCAGTGTTCGCTAGTCGATAT 324
  |||:::|||||

67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
  |||:::|||||
325 GGTAGATTCAAATAACAATTTAGAAGATAATAAGAACTTCCAAGGCGATGC 374
  |||:::|||||

84 laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
  |||:::|||||
375 CAGAGCATTCGCTCCCATTAAGATTCATCATTAATCAACGGGTAGAGAG 424
  |||:::|||||

101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTy 117
  |||:::|||||
425 TACCAATGGGATGCAGAAATCAAGGATGGATCTCGGTTCCTTTTATTA 474
  |||:::|||||

117 rGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134
  |||:::|||||
475 TGGGTACATCATCATCACAGATTCTCGGAGGATATGTGGCCAGCAAAATAG 524
  |||:::|||||

134 lyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
  |||:::|||||
525 GGGGGAAAATGCTGCTAGGATTTGGGATCCTTGGCACTGCTGTCTCCACC 574
  |||:::|||||

151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
  |||:::|||||
575 CTGTTCACATCCCATTTGCTGCAGATTTAGGAGTTGGACCACTCATTTGACT 624
  |||:::|||||

167 uArgAlaLeuGluClyLeuGlyGluGlyValThrTyrProAlaMetHisA 184
  |||:::|||||
625 CAGAGCACTAGAGGACTAGGAGGGGTGTACATTTTCAGCCATCCATCGAT 674
  |||:::|||||

184 laMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
  |||:::|||||
675 CCATGTGCTCTTCTTGGCTCCGCCCTCTTCAAGAAGCAAACTTCTTAGC 724
  |||:::|||||

```



```
259 rProTyrGluLysGluTyrIleLeuSerSerLeuLysAsnGlnLeuSerS 276
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
861 GAAGAGGAATGCCATACATCCAGAGACCATCGGTATCTGGGCTCTA 910
276 erGlnLysSerValProTrrPileProMetLeuLysSerLeuProLeuTrrp 292
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
911 AGCAGCTCAAGCACCCCTGGAGGGCCATCTTCACGTCGATGCCATTTAT 960
293 AlalIleValAlaAlaHisPheSerTyrAsnTrrPheTyrThrLeuLe 309
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
961 GCCATCATGGCCTCGCATTTCTCGGAGAACGGGGCTTCTACACCCGTCT 1010
309 uThrLeuLeuProThrTyrMetLysGluValLeuArgPheAsnIleGlnG 326
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1011 CACCCAGCTGCCAGCTTCTCCAGGACACGCTCAACTTTGACCTTGCCA 1060
326 luAsnGlyPheLeuSerAlaValProTrrLeuGlyCysTrrPLeuCysMet 342
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1061 AAACAGGCATTTCTTCGGCGGTTCCATACCTGGCCATGGCATCTGCTG 1110
343 IleLeuSerGlyGlnAlaAlaAspAsnLeuArgAlaArgTrpAsnPheSe 359
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1111 GCCGTGTCGGTATTGTCGTGATGGCTGCAGGTGAAGGCATTTGCAC 1160
359 rThrLeuTrrpValArgArgValPheSerLeuIleGlyMetIleGlyProA 376
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1161 GACCACCCAGGTTAGCGGGAACCTTCAACTGGGAGCCTTCTTGCCCCAGA 1210
376 laIlePheLeuValAlaAlaGlyPheIleGlyCysAspTyrSerLeuAla 392
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1211 CGGTGTTTCATGCTGACGCCCTATTG...CTGGATCCACGCTGGTCTG 1257
393 ValAlaPheLeuThrIleSerThrThrLeuGlyGlyPheCysSerSerG 409
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1258 GTGGTGAGTCTGACCATTCGGTGGGCTTGGGTGCCCTTCGCTGCTGG 1307
409 yPheSerIleAsnHisLeuAspIleAlaProSerTyrAlaGlyIleLeuL 426
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1308 ATTTCGGCTTAATCACCTCGACATTCGCCGCCACGACGCCAGCTGTTCTGA 1357
426 euGlyIleThrAsnThrPheAlaThrIleProGlyMetIleGlyProIle 442
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1358 TGGGCATTTGCCAACACATTTGCCACCATTCGGGCATCGTGAGTCCGCTG 1407
443 IleAlaArgSerLeuThrProGluAsnThrIleGlyGluTrrpGlnThrVa 459
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1408 CTACGGGCTACGTGGTCAACCAACACAGTAGTGACGAATGGCGCATAT 1457
459 lPheCysIleAlaAlaAlaIleAsnValPheGlyAlaIlePhePheThrL 476
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1458 CTCTTTCATTTCCGCGGCGCATTTACCTCGTGGCTGCGTCACTTTACTGGT 1507
476 euPheAlaLysGlyGluValGlnAsnTrrpAlaIleSerAspHisGln 491
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1508 TCTACTCTCGCGGATTTGCAGGAATGGGCCAAGACGCCGCGAGCAG 1554
```

seq_name: gb_htg:AC014246

```
seq_documentation_block:
LOCUS AC014246 33444 bp DNA linear HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC014246
VERSION AC014246.1 GI:6437089
KEYWORDS HTG; HGGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 33444)
AUTHORS Adams,M. and Venter,J.C.
```

TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDW:10212070 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES

source Location/Qualifiers

1..33444

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

BASE COUNT 9162 a 7329 c 7079 g 9874 t

ORIGIN

alignment_scores: Quality: 981.50 Length: 517

Ratio: 2.853 Gaps: 6

Percent Similarity: 66.538 Percent Identity: 38.685

alignment_block:

US-09-776-865-4 x AC014246 ..

Align seg 1/1 to: AC014246 from: 1 to: 33444

3 SerProValSerAspLeuAlaProSerAspGlyGluGluGlySerAspAr 19

||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

5996 TCCACTCTCTCTCTCCACCTCCAGAAACCATGGCTGTGGCCCAA 5945

19 qThrProLeuGlnArgAlaProArgAlaGluProAlaProValCysC 36

||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

5946 GACG..... 5949

36 ysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPhePheVal 52

||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

5950CGTCACATTTTCGGATTTCATGGGCTTCCTGGGATTCCGGGTG 5991

53 LeuTyrSerLeuArgValAsnLeuSerValAlaLeuValaspMetValas 69

:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

5992 GTCTACCGGATGCGGGTCAATCTCGTGGTGGCCATTGTGGCCATGGTG 6039

69 pSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysAlaGluH 86

||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

6040AACCAACGCAATTCGACACGACACTCATCGGTGATTGATACGG 6085

86 isSerAlaProIleLysValLeuHisAsn..... 95

:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

6086 ACAGTGTCCACTACCGGACCCACATCACAATGGTAGCATCCCAATCCG 6135

96 GlnThrGlyLysLysTyrArgTrrpAspAlaGluThrGlnGlyTrrpIleLe 112

||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

6136 CAGAAGGAGGCGGAGTTTGTGGGAGGAGCCACGAGGAGGATTGGTGCT 6185

112 uGlySerPhePheTyrGlyTrrpIleIleThrGlnIleProGlyGlyTrrv 129

:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

6186 CGGCAGTTTCTTCTATGGCTATGTGCTAACCAAGTGCCTGCCGCGGACGGA 6235

129 alAlaSerArgSerGlyGlyLysLeuLeuLeuGlyPheGlyIlePheAla 145

:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

6236 TGGCCGAGCTGTATGGTGGGAAGAAGATCTACGGCTATGGAGTGTGATC 6285

146 ThrAlaIlePheThrLeuPheThrProLeuAlaAlaaspPheGlyValG 162

||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

6286 ACGCGGTCTTTACGTATATAACTCCATTGGCTGCCCTGGGATGTCGCC 6335

162 yAlaLeuValAlaLeuArgAlaLeuGluGlyLeuGlyGluGlyValThr 179

||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

6336 GCTGTTGGTCTGTCGTCATCTCTGGAGGGAATGGGAGGCGGCTCACCT 6385

179 yrProAlaMetHisAlaMetTrrpSerSerTrrpAlaProProLeuGluArg 195

||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

6386 ATCCAGCTATGCACCCCAATGCTGCCACTGGATACCCCGCTGGAGAG 6435
196 SerLysLeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrValVa 212
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6436 AACAGTTGCGCGCAATCGTATGCGGGCTCCCAATATCGGAACAGTCAT 6485
212 IserLeuProLeuSerGlyValIleCys.....TyrTyrMetAsnT 226
:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6486 TTCCATGCGCTGGCGGATGCTGTGCTGCTGGACTTCCTGGGTGGCT 6535
226 rpThrTyrValPheTyrPheGlyIleValGlyIleIleTrpPheIle 242
||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6536 GCGCGTGGCTTCTACATCTTCGAGCTGTGGCATCTCTGTGGTTTCATC 6585
243 LeuTrpIleCysLeuValSerAspThrProGluThrHisLysThrIleTh 259
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6586 GCATGGATGTTATTTGGTGTACGAAGCTAGCGATCATCCAGGATCTC 6635
259 rProTyrGluLysGluTrpIleLeuSerSerLeuLysAsnGln..... 273
:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6636 AGAATCGGAGCGAGATATATCGAAGAGTCTACAGGTTGAGAGCTAA 6685
273 273
6686 TAAATCAGGATCTAGCGGAGCGGAGGAGGAGGAGGAGGATGAAGTG 6735
274LeuSerSerGlnLysSerValProTrpIleProMetLeuLy 287
6736 ACTTCGGGGGCGCGCGAGAGACCGATACCCCTGGTGCATCGTCTCAC 6785
287 sSerLeuProLeuTrpAlaIleValAlaHisPheSerTyrAsnTrpT 304
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6786 ATCCGTACTCTGTGGGCCATCTTTGTCACGCAATGCGCGCAGGATGGG 6835
304 hrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLysGluValLeu 320
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6836 CTTTCTACACGACGATACCCAGGTGTCCTACATGAGCAACATCCTA 6885
321 ArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValProTyrLeuG 337
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6886 CACTTGCATCCAGTCGAATGCTGTCTCAATGGGTGCGGTATCTAAC 6935
337 yCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAspAsnLeuArg 354
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6936 CTCCTGGTTCTGGCATTCGCTGCTCGCCCTGGCGGATGGATGCTAG 6985
354 laArgTrpAsnPheSerThrLeuTrpValArgArgValPheSerLeuIle 370
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
6986 CCAGCGCTACATATCGCTGCTGAACCTCGTATAAGTTGTGGAACCGGTG 7035
371 GlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyPheIleGlyCy 387
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7036 CCTTCGGTGGTCCATCACTGGGCTGATGGCATCATCTATGTGGGTG 7085
387 sAspTyrSerLeuAlaAlaPheLeuThrIleSerThrThrLeuGlyG 404
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7086 CGATTGGGTGGTGCACCTTATGTTGGCGCGGTGGCTCTTTCGGCG 7135
404 lPheCysSerSerGlyPheSerIleAsnHisLeuAspIleAlaProSer 420
||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7136 GCGCGCTCTATGCTGGCAACAGATGAATCATAGCGCTCAGTCCACGA 7185
421 TyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaThrIleProG 437
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7186 TATCGAGCCACCATGATGGCATCATCACCATTTCGGCGCAATATCTGTGG 7235
437 yMetIleGlyPro...IleIleAlaArgSerLeuThrProGluAsnThrI 453
||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7236 CTTCTGGCTCGGTATGTCATCGCTCTAAATCATCAATCATCGCGAGACTC 7285
453 leGlyGluTrpGlnThrValPheCysIleAlaAlaIleAsnValPhe 469
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7286 TGACACAGTGGCATCTGCTGCTGCTGGTGGCGGCGGCTTGAATATAGCC 7335

470 GlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnAsnTrpAl 486
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
7336 GGTAACCTTCATCTACCTCATCTTCCAGCCGCCGAGGAGCAAGCTGGTC 7385

486 a 486
7386 g 7386

seq_name: gb_in.AC023685

seq_documentation_block:

LOCUS AC023685 167928 bp DNA linear INV 03-JAN-2002
DEFINITION Drosophila melanogaster 3L BAC RP98-20N12 (Roswell Park Cancer
Institute Drosophila BAC Library) complete sequence.

ACCESSION AC023685
VERSION AC023685.3 GI:17223100
KEYWORDS HTG.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 167928)

AUTHORS

Gocayne, J.D., Tabor, P., Williamson, A., Homs, F.H.,
Dugan-Rocha, S.D., Sodergren, E.S., Hodgson, A.H., Chen, R.C.,
Ayale, M.A., Scott, G.S., Worley, K.W., Amaratunga, P.G., Brandon, R.C.,
Rogers, Y., An, H., Baldwin, D., Beeson, K.Y., Brown, M., Buhay, C.,
Busam, D.A., Center, A., Chen, G., Chen, Z., Clerc-Blankenburg, K.,
Davenport, L.B., Dietz, S.M., Ding, Y., Dodson, K., Doup, L.E.,
Draper, H., Emery-Cohen, A., Fertiera, S., Garg, N.D.S., Houck, J.,
Hostin, D., Howland, T.J., Hume, J., Ibegwam, C., Jalali, M., Kovar, C.,
Liu, W., Mattel, B., McIntosh, T.C., Morgan, M., Moy, M., Murphy, B.,
Nelson, K.A., Ndassa, Y., Nguyen, N., Perez, L., Pittman, G.S., Puri, V.,
Scheeler, F., Shen, H., Strong, R., Tector, C., Wang, Q., Williams, S.M.,
Xiang, J., Zaveri, J.S., Zhou, J., Zorrilla, S., Smith, H.O.,
Wheeler, D., Weinstock, G., Gibbs, R. and Venter, J.C.

Direct Submission

Unpublished

2 (bases 1 to 167928)

Worley, K.C.

Direct Submission

Submitted (17-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 167928)

Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbara, J.,
Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowle, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C.,
Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulsegh, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,

Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sison,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission

TITLE

Submitted (01-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

4 (bases 1 to 167928)
Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisgued,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sison,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission

TITLE

Submitted (01-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

5 (bases 1 to 167928)
Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisgued,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sison,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission

TITLE

Submitted (29-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

5 (bases 1 to 167928)
Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisgued,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sison,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission

TITLE

Submitted (29-DEC-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

5 (bases 1 to 167928)
Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisgued,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sison,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

alignment_scores:

| | | | |
|---------------------|--------|-------------------|--------|
| Quality: | 981.50 | Length: | 517 |
| Ratio: | 2.853 | Gaps: | 6 |
| Percent Similarity: | 66.538 | Percent Identity: | 38.685 |

alignment_block:

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Joudah,S., Karlsson,E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisgued,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sison,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission

TITLE

Submitted (01-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

AUTHORS

6 (bases 1 to 167928)
Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisgued,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N., Sison,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

REFERENCE
AUTHORS

1 (bases 1 to 174163)
Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,
Gocayne,J.D., Tabor,P., Williamson,A., Homs1,F.H.,
Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
Ayele,M.A., Scott,G.S., Worley,K.W., Amaratides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,
Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,
Davisport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
Draper,H., Emery-Cohen,A., Ferriera,S., Garg,N.D.S., Houck,J.,
Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,
Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,
Scheeler,F., Shen,H., Strong,N., Tector,C., Wang,Q., Williams,S.M.,
Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,
Wheeler,D., Weinstein,G., Gibbs,R. and Venter,J.C.
Direct Submission
Unpublished

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 174163)
Worley,K.C.
Direct Submission
Submitted (17-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS

3 (bases 1 to 174163)
Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Blimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgatz,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Homs1,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.B., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S.,
Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Submitted (19-DEC-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Dec 19, 2001 this sequence version replaced gi:6997282.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only

sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect <1e-34) to the
EST and cDNA sequences. Genes demonstrate at least 2 exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT.

| FEATURES | Source | Location/Qualifiers |
|--|---|---------------------|
| BASE COUNT | 51926 a 36149 c 35565 g 50523 t | |
| ORIGIN | | |
| alignment_scores: | Quality: 981.50 Length: 517 | |
| | Ratio: 2.853 Gaps: 6 | |
| | Percent Similarity: 66.538 Percent Identity: 38.685 | |
| alignment_block: | | |
| US-09-776-865-4 x AC023711 | | |
| Align seg 1/1 to: AC023711 from: 1 to: 174163 | | |
| 3 SerProValSerAspLeuAlaProSerAspGlyGluGluGlySerAspAr 19 | | |
| : : : : : : : | | |
| 120244 TCCACTCTCTCTCTCCACCCTCGCAGAAACCATGGCTGGCCCAA 120293 | | |
| 19 gThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProValCysC 36 | | |
| : : : : : : : | | |
| 120294 GAGC..... | | |
| 36 ysSerAlaArgTyAsnLeuAlaPheLeuSerPheGlyPheVal 52 | | |
| : : : : : : : | | |
| 120298CGTCACATTTTCGGATTCATGGCTTCCTGGGATTCGCCGTG 120339 | | |
| 53 LeuTySerLeuArgValAsnLeuSerValAlaLeuValAspMetValAs 69 | | |
| : : : : : : : | | |
| 120340 GTCTACGCGATCGGGTCAATCTCGGTGGCCATTTGGCCATGGTG.. 120387 | | |
| 69 pSerAsnThrThrAlaLysAspAsnArgThrSerTyGluCysAlaGluH 86 | | |
| : : : : : : : | | |
| 120388AACCAACGGCAATTCGGCAGACCACTCATCGGTGATTGATACGG 120433 | | |

```

86 isSerAlaProIleLysValLeuHisAsn..... 95
   :::::||||:   :::::||||:
120434 ACACGTGTCCACTACCGCGCACCATACATGGTAGGCATCCCAATCCG 120483

96 GlnThrGlyLysLysTyrArgTrpAspAlaGluThrGlnGlyTrpIleLe 112
   ||| ::::: ||||| ||||| ||||| ::||
120484 CAGAAGGAGGGCGAGTTGTGTGGACGAGGCCACGACGAGATTGGTGT 120533

112 uGlySerPhePheTyrGlyTyrIleIleThrGlnIleProGlyGlyTyr 129
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120534 CGCGAGTTCTTCTATAGCTATGTCTTAACCAAGTCGCCGCGGACGGA 120583

129 aLaSerArgSerGlyLysLeuLeuLeuGlyPheGlyIlePheAla 145
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120584 TGGCCGAGCTGTATGTTGGGAAGAAGATCTACGGCTATGGAGTGTGATC 120633

146 ThrAlaIlePheThrLeuPheThrProLeuAlaAlaAspPheGlyVal 162
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120634 ACGGCGGTCTTTACGCTTATTAACCTCAATGGCTGCCCACTGGGATCTGCC 120683

162 yAlaLeuValAlaLeuArgAlaLeuGluGlyLeuGlyGluGlyValThr 179
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120684 GCTGTGTGCTCTGTGCCCATCTCGAGGAATGGCGAGGCGTCACT 120733

179 yrProAlaMetHisAlaMetTrpSerSerTrpAlaProLeuGluArg 195
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120734 ATCCAGCTATGCACGCCATGCTTGCCCACTGGATACCGCCGCTGGAGG 120783

196 SerLysLeuLeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrVal 212
   ::||| ::||| ||||| ||||| ||||| ||||| ||||| |||||
120784 AACAAAGTTTCGGCGCAATCGTATATCGGGGCTCCAATATCGGAACAGTCA 120833

212 lSerLeuProLeuSerGlyValIleCys.....TyrTyrMetAsn 226
   :|||: ||||| ||||| ||||| ||||| ||||| ||||| |||||
120834 TTCCATGCGCTGGCCGAGTGGTGTCGCTGCGTGGACTTCTGGTGGCT 120883

226 rpThrTyrValPheTyrPheGlyIleValGlyIleIleTrpPheIle 242
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120884 GGCCGTCGGCTTTCTACATCTTCGACGTGTGGCATCTCTGGTTCATC 120933

243 LeuTrpIleCysLeuValSerAspThrProGluThrHisLysThrIleTh 259
   ||||| ::||| ||||| ||||| ||||| ||||| ||||| |||||
120934 GCATGGATGTTATTTGGTGTACGACAAAGCTACCGATCATCCAGGATCTC 120983

259 rProTyrGluLysGluTyrIleLeuSerSerLysLeuAsnGln..... 273
   : ||||| ||||| ||||| ||||| ||||| ||||| |||||
120984 AGAATCGGCGGAGAGTATATCGAAAGAGTCTACAGGTTGAGGCTAA 121033

273 ..... 273
121034 TAAATCAGGATCTAGCGGAGCGCGAGGAAGAGGAGGACGATGAAGTG 121083

274 .....LeuSerSerGlnLysSerValProTrpIleProMetLeuL 287
   ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
121084 AGTCTGGCGGCGCGCGGAGGAACCGATACCCCTGGTCTGCTCAC 121133

287 sSerLeuProLeuTrpAlaIleValValAlaHisPheSerTyrAsnTrp 304
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121134 ATCCGATCTCTGTGGGCCATCTTGTGACCAATCGCGCAGGATGGG 121183

304 hrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLysGluValLeu 320
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121184 CCTTCTACACGACGATACCGAGCTGCCACCTACATGAGCAACATCCTA 121233

321 ArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValProTyrLeuGl 337
   ::||| ::||| ||||| ||||| ||||| ||||| ||||| |||||
121234 CACTTTGACATCCAGTCCGAATGCTGTGCTCAATCGGTGCGGTACTAAC 121283

337 yCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAspAsnLeuArg 354
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121284 CTCCTGGTTCGTGGCATGTGCTGCTCGGCCCTCGCGGATGGATGCTAG 121333

354 laArgTrpAsnPheSerThrLeuTrpValArgValPheSerLeuIle 370

```

```

||||| ::||| ||||| :::::||||:   :::::||||:   ::||
121334 CCAGACGCTACATATCGCTGAACCTCGTATAGTTGTGGAACACGGTG 121383

371 GlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyPheIleGlyCy 387
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121384 GCCTCGCGTGGTGCATCATCGGCTGATTGGCATCATCTATGTGGGCTG 121433

387 sAspTyrSerLeuAlaValAlaPheLeuThrIleSerThrThrLeuGlyG 404
   ||||| ::||| ||||| ||||| ||||| ||||| ||||| |||||
121434 CGATTGGTGTGGGTACCTTTATGTGGCGCGGTGGGCTCTTTTCGGCG 121483

404 lypheCysSerSerGlyPheSerIleAsnHisLeuAspIleAlaProSer 420
   || ::||| ::||| ||||| ||||| ||||| ||||| |||||
121484 CGCGCGCTATATGTCGCAACAGATGAATCATAGCGCTCAGTCCACGA 121533

421 TyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaThrIleProGl 437
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121534 TATCAGGACCATGTATGGCATCACCAAATTCGGCGGCAATATCTGTGG 121583

437 yMetIleGlyPro...IleIleAlaArgSerLeuThrProGluAsnThrI 453
   ::||| ::||| ||||| ||||| ||||| ||||| ||||| |||||
121584 CTTCTGCTCCGTATGTCATCGGTCTTAATCATCAATCATCGCGAGACTC 121633

453 leGlyGluTrpGlnThrValPheCysIleAlaAlaIleAsnValPhe 469
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121634 TGACACAGTGCATCTGTGCTCTTCGTGGCGCGGCTTGAATATAGCC 121683

470 GlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnAsnTrpAl 486
   ||| ::||| ::||| ||||| ||||| ||||| ||||| |||||
121684 GGTAACTTCACTACCTGATCTGCCAGCGCGGAGGCAAGCTGCTC 121733

486 a 486
121734 G 121734

seq_name: gb_in:AE003491
seq_documentation_block:
LOCUS AE003491
DEFINITION Drosophila melanogaster genomic scaffold 142000013386053 section 8
of 30, complete sequence.
ACCESSION AE003491 AE002593
VERSION AE003491.2 GI:10728219
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 323461)
AUTHORS Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Aghayani,A., An,H.J.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkov,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Doddson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.,

```


1547 GAGGAAGTTTTCACATCCATCCCGCTACGCGATAATTTGTCGCAACT 1596
 299 heSerTyAsnTrpThrPheTyThrLeuLeuThrLeuLeuProThrTy 315
 1597 TCTGCAGAGCTGACCTTTTATTACTTGCTCATCAGTCACGCCAGCTTAT 1646
 316 MetLysGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAl 332
 1647 TTTGAGGAGGTTTGGATTGAAATCAGCAAGGTTGGCATGTTGCTGCG 1696
 332 aValProTyLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaA 349
 1697 AGTCCCTCACCTTCATGACAAATCATTTGCTTCATCGGGGGGCAATTG 1746
 349 laAspAsnLeuArgAlaArgTrpAsnPheSerThrLeuTrpValArgArg 365
 1747 CAGATTCTTAGAGCAGCAAGAAATTCCTCAACAACACTACAGTGAGAAG 1796
 366 ValPheSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAl 382
 1797 ATCATGAATCTGGGGTTTGGCATGGAAGCCAGCGCTGCTTCTGGTTGT 1846
 382 aGlyPheIleGlyCysAspTySerLeuAlaValAlaPheLeuThrIleS 399
 1847 TGGTAC...TCTCATACTAGAGGGTGGCCATCTCTTCTTGGTGTG 1893
 399 erThrThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeu 415
 1894 CAGTAGGATTCAGTGATTCTCTATCTGTTTCAATGTTAATCACTTG 1943
 416 AspIleAlaProSerTyAlaGlyIleLeuLeuGlyIleThrAsnThrPh 432
 1944 GATATGCTCCAGATATGCCAGTATCTTATGGCATTTCAATGGCGT 1993
 432 eAlaThrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrP 449
 1994 TGGACGCTGCGGGATGTTGTCCTCATCATTTGTTGGTCAATGACAA 2043
 449 roGluAsnThrIleGlyGlnTrpGlnThrValPheCysIleAlaAlaAla 465
 2044 AGATAGTCCGTGAGATGAGATGATGCTTCTTCATTTGCTGCACATC 2093
 466 IleAsnValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluVa 482
 2094 GTCCACTATGCTGGAGTCATATTTATGCACTATTGCTCCTCAGGAGAA 2143
 482 lGlnAsnTrpAla 486
 2144 ACAACCTTGGGCA 2156

seq_name: gb_ro:AF271235

seq_documentation_block:
 LOCUS AF271235 3982 bp mRNA linear ROD 14-JUN-2000
 DEFINITION Rattus norvegicus differentiation-associated Na-dependent inorganic phosphate cotransporter (DNPI) mRNA, complete cds.

ACCESSION AF271235
 VERSION AF271235.1 GI:8515880

KEYWORDS Norway rat.
 SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 3982)
 AUTHORS Mashima,H. and Kojima,I.

TITLE Rat DNPI
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3982)
 AUTHORS Mashima,H. and Kojima,I.

TITLE Direct Submission
 JOURNAL Submitted (24-MAR-2000) Cell bilology, Gunma University, Institute for Molecular and cellular Regulation, 3-39-15, Showa-machi,

FEATURES
 Location/Qualifiers
 source 1..3982
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 gene 1..3982
 /gene="DNPI"
 CDS 463..2211
 /gene="DNPI"
 /codon_start=1
 /product="differentiation-associated Na-dependent inorganic phosphate cotransporter"
 /protein_id="AAF76223.1"
 /db_xref="GI:8515881"
 /translation="MESVKORILAPKGEIKNPACKSLGOIYRVLEKKODNBETIELT EDGKLEPEKRAPLCDCTCGLPRIYIINSGLFCISFGIRNGLNGLVAIVDMVNNST THRGKGVIKERAKFNWDPEYVMIHGSEFFWYIIITQIPGGYIARLNARVFGAAIL LTNLMILPSAARHYGVIEFVRLQGLVEGYTPACHGWSKAPPLERSLRATTS FCGSYAGAIAMPLAGILVQYVGSFVYVGSFGVMWYMWLLVSYSPAKHPHTTD BERRYIEESIGESANLIGAMEKFKTPRKFFTSMPYATIVANFCRSTFYLLLSQP AYEEVEGFEISKVGLMSAVPHLVMTILVPIGGIADFLRSKOILSTTVTKIMCGG FGMETALLVGVYSHTRGVAISFLVAVGFSFAGSGFNVNHLDIAPRYASILMGISN GVTLSGMVCPILIVGAMTKNKRSEWQYVFLIALVHVGVIIFYALFASGEKOPWADP EETSEKCGFIHDEDEETGDTQNYINYGTTKSYGATSQENGWPNWGEKKEEFVQ ESAQDAYSYKDRDDYS"
 poly_a_signal 3953..3958
 /gene="DNPI"
 BASE COUNT 1144 a 823 c 884 g 1131 t
 ORIGIN

alignment_scores:

Quality: 968.00 Length: 502
 Ratio: 2.742 Gaps: 8
 Percent Similarity: 70.319 Percent Identity: 40.637

alignment_block:

US-09-776-865-4 x AF271235 ..

Align seg 1/1 to: AF271235 from: 1 to: 3982

2 LysSerProValSerAspLeuAlaProSerAspGlyGluGlySerAs 18
 525 AAAATCCCTCGGACAGATCTACAGGGTGCTGGAGAAGAAGCAGGATAACC 574
 18 pArgThrProLeuLeuGlnArgAlaProArgAla.....G 30
 575 GAGAGACCATCGAGCTCACAGAGGACGGCAAGCCCTGGAGGTGCCTGAG 624
 30 lu.ProAlaProValCys...CysSerAla.....ArgTy 40
 625 AAGAAGGCTCGGTATCGGACTGTACGTGCTTCGGCCTCGCGCCGCTA 674
 40 rAsnLeuAlaPheLeuSerPheGlyPhePheValLeuTySerLeuA 57
 675 CATCATAGCCATCATGACGGCCTCGGCTTCTGCATCTCCTTTGGTATCC 724
 57 rgValAsnLeuSerValAlaLeuValAspMetValAspSerAsnThr 73
 725 GCTGTAACCTGGGTGGCCATTGTGGACATGGTCAACAACAGCACCATC 774
 74 AlAlaYAspAsnArgThrSerTyrgLysAlaGluHisSerAlaProI 90
 775 CACCGGGGTGCACCAAGTTATCAAGGAGAAAGCC..... 807
 90 eLysValLeuHisAsnGlnThrGlyLysLysTyArgTrpAspAlaGlu 107
 808AAGTTAACTGGGACCCGAGA 829
 107 hrGlnGlyTrpIleLeuGlySerPhePheTyrgLlyTyrlleThrGln 123
 830 CTGTGGGGGATGATTCACGGGTGCTTCTCTGGGGCTATATCATCACCAG 879


```

GEVQDSHSYKDRVDYS"
polyA_signal 3895..3900
polyA_site 3921
BASE COUNT 1238 a 735 c 791 g 1182 t
ORIGIN
alignment_scores:
  Quality: 963.50      Length: 492
  Ratio: 2.737        Gaps: 9
  Percent Similarity: 71.545      Percent Identity: 41.057
alignment_block:
US-09-776-865-4 x AB032435 ..
Align seg 1/1 to: AB032435 from: 1 to: 3946
13 GlyGluGlu.....GlySerAspArgThrProLeuLeuGlnAr 25
|||||
506 GGAGAAGAAGCAACACCGGGGAGACAATCGAGCTGACGGAGGATGGGA 555
25 gAlaProArgAla.....Glu..ProAlaProValCys...CysSerAla 38
|||||
556 AGCCCCCTAGAGGTGCCCGAGAGAGAGCGCGCTGTGCGACTGCACGTCG 605
39 .....ArgTyrAspLeuAlaPheLeuSerPhePheGlyPh 50
|||||
606 TTCGCCCTGCCCGCCCTACATTTATGCCCATCATGAGCGCCCTGGGCTT 655
50 ePheValLeuTyrSerLeuArgValAlaLeuSerValAlaLeuValAspM 67
| : : : : :
856 CTGCATCTCCTTCGTATCCGCTCAACCTGGCGGTGGCCATTGTGGACA 705
67 etValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCys 83
| : : : : :
706 TGGTCAACACAGACACCATCCACCGCGGGGCAAGGTCAACAGGAGAAA 755
84 AlaGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLy 100
|||
756 GCC.....AA 760
100 sTyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePhe 117
| : : : : :
761 ATTCAACTGGGACCGGAAACCGTGGGATGATCCACGGTTCCTCTTT 810
117 yrGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSer 133
| : : : : :
811 GGGGCTACATCATCATCAGATTCGGGAGGCTACATCGCGCTCTCGGCTG 860
134 GlyGlyLysLeuLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheTh 150
| : : : : :
861 GCACCCAAACAGGGTTTCGGAGCTGCCATACCTTTACCTTACCCCTAAA 910
150 rLeuPheThrProLeuAlaAspPheGlyValGlyAlaLeuValAlaL 167
| : : : : :
911 TATGCTAATTCATCAGCAGCAGCAGTGATGATGATGATGATGATGATG 960
167 euArgAlaLeuGluGlyLeuGlyGlyValThrTyrProAlaMetHis 183
| : : : : :
961 TCAGAACTACGACGGGACTTGTTCAGGGGTGTGACCTACCCAGCATGTCT 1010
184 AlaMetTrpSerTrpAlaProProLeuGluArgSerLysLeuLeuSe 200
| : : : : :
1011 GGGATATGGAGCAATGGGCCCCACCTCTAGAGAGGAGTAGACTGGCAAC 1060
200 rIleSerTyrAlaGluGlnLeuGlyThrValValSerLeuProLeus 217
| : : : : :
1061 CACCTCCTTTTGTGCTTCCATGCGGAGCTGTGATTGCAATGCTTTAG 1110
217 erGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePhe 233
| : : : : :
1111 CTGGCATCTTGTGCGACTACACTGGCTGGTCTTCAGTGTTTTATGTCTAC 1160

```

```

234 GlyIleValGlyIleIleIleTrpPheIleLeuTrpIleCysLeuValSerAs 250
|||
1161 GGAAGCITTGGAATGGCTCGTACATGTTTGGCTTTGGTGTCTTATGA 1210
250 pThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleL 267
| : : : : :
1211 AAGTCCTGCAAGCATCTCTACTATTACAGATGAAGACGCTAGGTACATAG 1260
267 euSerSerLeu.....LysAsnGlnLeuSerSerGlnLys..... 278
| : : : : :
1261 AAGAAAGCATTTGGAGAGAGTGCAAAATCTTTAGTGTGCAATGGAAAAATTC 1310
279 SerValProTrpIleProMetLeuLysSerLeuProLeuTrpAlaIleVa 295
| : : : : :
1311 AAGACTCCATGGGAGGAAGTTTTCATCCATCCAGTCAGTCATGCAATAAT 1360
295 lValAlaHisPheSerTyrAsnTrpThrPheTyrThrLeuLeuThrLeuL 312
| : : : : :
1361 TGTTCGAAACTTCGCAAGAGCTGGACTTTTATTATTATTGCTTATTAGTC 1410
312 euProThrTyrMetLysGluValLeuArgPheAsnIleGlnGluAsnGly 328
| : : : : :
1411 AGCCAGCATATTTTGAGGAAGTCTTTGGATTGAAATTAGCAAGGTGGT 1460
329 PheLeuSerAlaValProTyrLeuGlyCysTrpLeuCysMetIleLeuSe 345
| : : : : :
1461 ATGCTATCTGCTGGCCACACTTAGTAATGACAATTTATGTGCCTATTGG 1510
345 rGlyGlnAlaAlaAspAsnLeuArgAlaArGTrpAsnPheSerThrLeuT 362
| : : : : :
1511 GGGCAAAATTCAGATTTTTCAGAAAGCAAGCAGATCTTTTCAACACTACA 1560
362 rpValArgArgValPheSerLeuIleGlyMetIleGlyProAlaIlePhe 378
| : : : : :
1561 CAGTGAGAAAGATCATGAATTTGGTGGTTTGGCATGGGAAGCCACACTG 1610
379 LeuValAlaAlaGlyPheIleGlyCysAspTyrSerLeuAlaValAlaPh 395
| : : : : :
1611 CTCTGTGCTGGTGGCTAT...TCTCATACTAGAGGGGTAGCAATCTCAT 1657
395 eLeuThrIleSerThrThrLeuGlyGlyPheCysSerSerGlyPheSerI 412
| : : : : :
1658 CTTGGTACTTGCAGTGGGATTCAGTGGATTGCTATATCTGTTTCAATG 1707
412 leAsnHisLeuAspIleAlaProSerTyrAlaGlyIleLeuLeuGlyIle 428
| : : : : :
1708 TTAACCACTTGGATATCGTCCAAAGATATGCCAGTATCTTAAATGGGCT 1757
429 ThrAsnThrPheAlaThrIleProGlyMetIleGlyProIleIleAlaAr 445
| : : : : :
1758 TCGAATGGTGTGGCAGATGTCAGGAATGGTTTGTCTCTATCATTTGTTG 1807
445 gSerLeuThrProGluAsnThrIleGlyGluTrpGlnThrValPheCysI 462
| : : : : :
1808 TGCAATCACAAAGAAATAAGTCACGCTGAAGAGTGGCAGTATGCTCTCTGA 1857
462 leAlaAlaAlaIleAsnValPheGlyAlaIlePhePheThrLeuPheAla 478
| : : : : :
1858 TCGCTGCCCTAGTCCACTATGTTGGAGTTATATTTTATGCAATATTGGC 1907
479 LysGlyGluValGlnAsnTrpAla 486
| : : : : :
1908 TCAGGAGAGAAACAAACCTGGGCA 1931

```


THIS PAGE BLANK (USPTO)

alignment_block:

US-09-776-865-4 x AAZ50876 ..

Align seg 1/1 to: AAZ50876 from: 1 to: 2844

```
1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGluGlySe 17
|||||
84 ATGAAGTCCCGGTTTCGGACTTAGCCCGAGCGAGCGGAGGGGCTC 133
|||||
17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProV 34
|||||
134 GAACCGCACACCGCTCCCTGCAGCGCGCCCGGGCGGAACCGCTCCAG 183
|||||
34 aLysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPhe 50
|||||
184 TATGCTGCTCTGCTCGTTACAACTAGACATTTTGTCTTTTGGTTTC 233
|||||
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
|||||
234 TTCGTTCTCTATTCAATTACGGGTGAATCTGAGCGTTGCACCTAGTGGACAT 283
|||||
67 tValAspSerAsnThrAlaLysAspAsnArgThrSerTyrGluCysA 84
|||||
284 GGTGGATTCAAAACACAACTGCCAAAGATAATAGAACGCTCTACGAGTGTG 333
|||||
84 LaGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLys 100
|||||
334 CAGAGCATCTGCTCCCAATAAAGTCTTCACAAACCAACGGGTAAAGAAG 383
|||||
101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTy 117
|||||
384 TACCGGTGGGATGCAGAAACTCAAGGATGGATTCTCGGATCTTTTCTA 433
|||||
117 rGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134
|||||
434 TGGCTACATCATCACAAATTCCTGGAGGATATGTTGCCAGCAGAAAGTG 483
|||||
134 LyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
|||||
484 GGGGAACCTGTTGCTAGGANTCGGGATCTTTGCTACAGCTATCTTCACC 533
|||||
151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
|||||
534 CTGTTCACTCCCTCGCTGCAGATTTCGGAGTCGGAGCCCTTGTGCACT 583
|||||
167 uArgAlaLeuGluGlyLeuGlyGluGlyValThrTyrProAlaMethHisA 184
|||||
584 CAGGGCACTAGAAGGGCTTAGGAGAGGGGTGCACATATCCAGCCATGCATG 633
|||||
184 LaMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
|||||
634 CCATGTGCTCTTCATGGGCTCCCCCTTTGAAGAAGCAAGCTTCTGAGT 683
|||||
201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
|||||
684 ATTTTCATATGCAGGACACAACTTGGGACAGTACTTCTCTCTCTCTTC 733
|||||
217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
|||||
734 TGGAGTAATTTGCTACTATATATGAAATTGGACTTATGTCTTCTATTCTTTG 783
|||||
234 LyIleValGlyIleIleThrPheIleLeuTrpIleCysLeuValSerAsp 250
|||||
784 GCATGTTGGAAATCATCTGGTTATTTATGGATCTGCTTAGTTAGTGTAT 833
|||||
251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
|||||
834 ACACCAGAACTCACAGACAATCACTCCCTATCGAAAAGGAGTATATTCT 883
|||||
267 uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284
|||||
884 TTTCATCATTAATAAATCAGCTCTCTTCACAGAAAGTCAGTCCGCTGGATAC 933
|||||
```

```
284 roMetLeuLysSerLeuProLeuTrpAlaIleValValAlaHisPheSer 300
|||||
934 CTATGCTGAATCACTGCCACTTTGGGCTATTGCTGTGCACATTTTCT 983
|||||
301 TyrAsnTrpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLy 317
|||||
984 TACAACCTGGACTTTTATACCTTTGTTGACCTTATTTGCTTACTTACATGAA 1033
|||||
317 sGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValP 334
|||||
1034 GGAAGTCTCAAGGTCAATATTCAGAGATGGGTTTTTATCTGCAGTCC 1083
|||||
334 roTyrLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 350
|||||
1084 CTTATTAGGTTGTTGGTTATGATGATCCGTGGGTCAAGCTGCTGAC 1133
|||||
351 AsnLeuArgAlaArgTrpAsnPheSerThrLeuTrpValArgArgValPh 367
|||||
1134 AATTTAAGGGCAAGATGGAATTTTCAACTCTGTGGGTTCGAAGAGTTTT 1183
|||||
367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaLaGlyP 384
|||||
1184 TAGCCTTATAGGGATGATTGGACCTGGCATATTCTGTTGCCGAGGAT 1233
|||||
384 heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
|||||
1234 TTATAGGCTGTGATATTCTCTGGCTGTGCTTCCCTAACCATATCAACA 1283
|||||
401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIl 417
|||||
1284 ACCCTGGAGGCTTTTGTCTCTCTGGATTTTAGCATCAACCATCTGGACAT 1333
|||||
417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaT 434
|||||
1334 TGCTCTCTCGTATGCTGTGTTATCTCTGGCATCACAAATACCTTGGCA 1383
|||||
434 hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
|||||
1384 CTATTCTCGNAATGATTGGGCCCAATCATGTCAGAAAGTCTTACCCCTGAG 1433
|||||
451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs 467
|||||
1434 AACACTATTGGAGAAATGGCAAACTGTTTCTGTCATCGCTGCTATCAA 1483
|||||
467 nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnA 484
|||||
1484 TGTATTTGGTGCCATTTTCTTCACACTATTGCCAAAAGGTGAAGTGCAA 1533
|||||
484 snTrpAlaIleSerAspHisGlnGlyHisArgAsn 495
|||||
1534 ACTGGGCCATCAGTGCATCACCAAGACACAGAAAC 1568
|||||
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAD10326

seq_documentation_block:

ID AAD10326 standard; DNA; 2844 BP.

XX AAD10326;

XX

DT 16-OCT-2001 (first entry)

XX

DE Sheep group B beta-haemolytic Streptococci toxin receptor (SP55) DNA.

XX

KW Sheep; group B beta-haemolytic Streptococci toxin receptor; GBS; SP55;

KW cytostatic; vulnery; antiatherosclerotic; osteopathic; vasotropic;

KW prevention; attenuation; pathogenic condition; cancer; scar;

KW wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;

KW keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;

KW vaccine; ds.

XX

OS Ovis sp.

XX

FH Key Location/Qualifiers
 FT CDS 84..1571
 FT /*tag= a
 FT /product= "Sheep GBS toxin receptor protein (SP55)"
 XX WO200156598-A2.
 PN 09-AUG-2001.
 XX 02-FEB-2001; 2001WO-US03662.
 XX 02-FEB-2000; 2000US-0179870.
 PR (UYVA-) UNIV VANDERBILT.
 XX Hellerqvist CG;
 XX WPI: 2001-488844/53.
 DR P-PSDB; AAE06519.
 XX Preventing or attenuating pathoangiogenic conditions e.g. cancer,
 PT chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by
 PT administering group B beta-hemolytic Streptococci toxin receptor or its
 PT fragment
 XX Disclosure: Page 47-50; 52pp; English.
 XX The present sequence is a DNA encoding sheep group B beta-haemolytic
 CC Streptococci (GBS) toxin receptor protein, SP55. The present invention
 CC relates to a method for preventing or attenuating a patho-angiogenic
 CC condition in a mammal which comprises administering to the mammal one
 CC or more GBS toxin receptors or their immunogenic fragments to induce
 CC or maintain an immune response to one of GBS toxin receptors. The
 CC method is useful for preventing or ameliorating pathoangiogenic
 CC conditions such as cancer, scarring during wound healing, gliosis
 CC during repair of nerve injury, chronic wounds, keloids, reperfusion
 CC injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and
 CC psoriasis in mammals. The proteins of the invention are also used
 CC as vaccines.
 XX
 SQ Sequence 2844 BP; 745 A; 623 C; 594 G; 882 T; 0 other;

alignment_scores:

Quality: 2617.00 Length: 495
 Ratio: 5.287 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-776-865-4 x AAD10326 ..

Align seg 1/1 to: AAD10326 from: 1 to: 2844

1 MetLysSerProValSerAspLeuAlaProSerAspGluGluGlySe 17
 CC |||||
 84 ATGAAGTCCCGGTTTCGGACTTAGCCCCGAGCGACGGCAGGAGGGCTC 133
 17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProv 34
 CC |||||
 134 GGACCGCACACCGCTCTGTCAGCGCGCGCCGCGGCGGAACCGCTCCAG 183
 34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPhe 50
 CC |||||
 184 TATGCTGCTCTGCTCGTTACAACCTAGCATTTTGTCTTTTGGTTTC 233
 51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
 CC |||||
 234 TTCGTTCTCTATTCAATACGGGTGAATCTGAGCGTTTGCACGTAGTGACAT 283
 67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
 CC |||||
 284 GGTGGATTCAACACACACTCCCAAGATAATAGAACGTCCTACGAGGTG 333

84 laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
 CC |||||
 334 CAGAGCATTCGTCCCAATAAAGTTCTTCAACACCAACCGGTAAGG 383
 101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTy 117
 CC |||||
 384 TACCGGTGGGATCGAGAACTCAAGGATGGATTCTCGGATCTTTTCTA 433
 117 rGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134
 CC |||||
 434 TGGCTACATCATCACAAATTCCTGGAGGATATGTTGCCAGAGAAGTG 483
 134 lyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
 CC |||||
 484 GGGGAAGCTGTTGCTAGGATTCGGGATCTTGTCTACAGCTATCTTCACC 533
 151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
 CC |||||
 534 CTGTTCACTCCCTCGCTCGAGATTCGGAGTCGGAGCCCTTGTTCACCT 583
 167 uArgAlaLeuGluGlyLeuGlyGluGlyValThrTyrProAlaMethHis 184
 CC |||||
 584 CAGGGCACTAGAAGGGCTAGGAGGGGTGTACATATCCAGCCATGCATG 633
 184 laMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
 CC |||||
 634 CCATGTGCTTCTATGGGCTCCCTCTTGAAGAGCAAGCTTCTGAT 683
 201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
 CC |||||
 684 ATTTCATATCAGGAGCACAACTTGGGACAGTAGTTCTCTCTCTCTTTC 733
 217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
 CC |||||
 734 TGGAGTAATTTGCTACTATATGAATGGACTATGCTCTTCTATTTCTT 783
 234 lyIleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
 CC |||||
 784 GCATTGTTGGAATCATCTGGTTATTTTATGGATCTGCTTAGTTAGTAT 833
 251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
 CC |||||
 834 ACACGAGAACTCACAAGACAATCACCCTCCCTATGAAAGGAGTATATCT 883
 267 userSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284
 CC |||||
 884 TTCATCATTAATAAATCAGCTCTCTTCAAGAGTCAAGTCCCGTGGATAC 933
 284 roMetLeuLysSerLeuProLeuTrpAlaIleValValAlaHisPheSer 300
 CC |||||
 934 CTATGCTGAAATCACTGCCACTTTGGGCTATTTGCTGTTGCACATTTTCT 983
 301 TyrAsnTrpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLy 317
 CC |||||
 984 TACAACCTGGACTTTTATACCTTTGTTGACCTTATTCCTACTTACATGAA 1033
 317 sGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValP 334
 CC |||||
 1034 GGAAGTCCCTAAGTTCAATATTCAAGAGAATGGGTTTTTATCTGCACCTC 1083
 334 roTyrLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 350
 CC |||||
 1084 CTTATTTAGGTTGTTGTTATGATGATCTCTGCGGTCAAGCTGCTGAC 1133
 351 AsnLeuArgAlaArgTrpAsnPheSerThrLeuTrpValArgArgValPh 367
 CC |||||
 1134 AATTAAAGGCAAGATGGAATTTTCAACTCTGTGGGTTCCGAGAGTTT 1183
 367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
 CC |||||
 1184 TAGCCTTATAGGATGATTTGGACCTGCGATATTCCTGTTGCCCGCAGAT 1233
 384 heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400

```
|||||TTATAGGCTGTGATTATTCCTTGGCTGTGCAATTCCTAAACCATATCAACA 1283
401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIle 417
|||||ACCTGGAGGCTTTTGGCTCTTCTGGATTTAGCATCAACCATCTGGACAT 1333
417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaIat 434
|||||TGCCTTCGTATCTGCTATCTCTCTGGGCATCACAAATACCTTTGCCA 1383
434 hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
|||||CTATTCCTGGAATGATGGGCCCATCATTCGCCAAGCTTTACCCCTGAG 1433
451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs 467
|||||AACACTATTGAGAATGCAAACTGTTTCTGCTATCCTGCTGCTATCAA 1483
467 nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnA 484
|||||TGTATTGGTGCCATTCTTCACACATATTGCCCAAGGTGAAGTGCAAA 1533
484 snTrpAlaIleSerAspHisGlnGlyHisArgAsn 495
|||||ACTGGCCATCAGTGATCACCAAGGACACAGAAAC 1568
```

seq_name: /SDSI/gc9data/hold-geneseq/geneseqn-emb1/WA2001A.DAT:AAF55900

```
seq_documentation_block:
ID   AAF55900 standard; DNA; 2512 BP.
XX
AC   AAF55900;
XX
DT   18-APR-2001 (first entry)
XX
DE   Human AST coding sequence.
XX
KW   Human; AST; nootropic; immunotropic; gene therapy; Salla disease;
KW   anion and sugar transporter; anion-cation symporter;
KW   sialic acid transporter; ss.
XX
OS   Homo sapiens.
PN   EP1069184-Al.
XX
PD   17-JAN-2001.
XX
PF   16-JUL-1999; 99EP-0202341.
XX
PR   16-JUL-1999; 99EP-0202341.
XX
PA   (ALKU ) AKZO NOBEL NV.
XX
WPI; 2001-193090/20.
DR   P-PSDB; AAB66967.
XX
PT   New human transporter gene implicated in Salla disease and lysosomal
PT   sialic acid transport, useful in assays for identifying new drugs, or
PT   diagnosing sialic acid transport defects related to mutations in the
PT   transporter gene
XX
PS   Claim 2; Page 12-13; 20pp; English.
XX
CC   The present sequence is the coding sequence for human Anion and Sugar
CC   Transporter (AST) protein. AST has significant homology with several
CC   members of the anion-cation symporter (ACS) family of transporters. AST
CC   is implicated in Salla disease, and is useful in screening assays for
CC   identifying new drugs. Compounds identified via AST screening is useful
CC   for preparing a pharmaceutical suitable as an activator or inhibitor of a
CC   sialic acid transporter protein. The pharmaceutical may be used in sialic
CC   acid associated diseases and CNS/immune related disorders.
XX
```

```
SQ   Sequence 2512 BP; 612 A; 585 C; 566 G; 749 T; 0 other;

alignment_scores:
    Quality: 2329.00      Length: 495
    Ratio: 4.903         Gaps: 0
    Percent Similarity: 95.960    Percent Identity: 86.263

alignment_block:
US-09-776-865-4 x AAF55900  ..
Align seg 1/1 to: AAF55900 from: 1 to: 2512
1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGluGlySe 17
|||||ATGAGGTCTCCGGTTCGAGACCTGGCCCGGAACGATGGCGAGAGACAC 320
17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaPro 34
|||||GGACCGCACGCCCTCTCTACGGGCGGCCACGGCGCGAGCCGCTCCAG 370
321 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPhe 50
|||||TGTGCTGCTCTGCTCGTTACAACTTAGCAATTTTGGCTTTTGGTTTC 420
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
|||||TTCATTGTGATGATTACGTGTGAATCTGAGTGTTCGTTAGTGGATAT 470
67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
|||||GGTAGATTCAAAATACAACCTTAGAAGATATAGAACTTCCAAGCGCTGC 520
84 laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
|||||CAGAGCATTTCTCCCATAAAGTTTCATATAATCAAGGGGTAGAAG 570
101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTy 117
|||||TACCAATGGGATGAGAACTCAAGGATGGATTCCTGGTTCCTTTTATA 620
117 rGlyTyrIleIleThrGlnIleProGlyTyrValAlaSerArgSerG 134
|||||TGGCTACATCATCACACAGATTCTCTGGAGGATATGTTGCCAGCAAAATAG 670
134 lyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
|||||GGGGGAAATGCTGCTAGGATTTGGGATCCTTGGCACTGCTGCTCACC 720
151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
|||||CTGTTCACTCCCAFTGCTGAGATTTAGGAGTTGGACCACCTCAFTGTACT 770
167 uArgAlaLeuGluGlyLeuGlyGluGlyValThrTyrProAlaMethHis 184
|||||CAGAGCACTAGAAGGACTAGGAGAGGGTGTACATTTCCAGCCATGCATG 820
184 laMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
|||||CCATGTGCTCTCTTGGGCTCCCCCTCTTGAAGAGCAAACTCTTTAGC 870
201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
|||||ATTTCATATCGAGGACAGCTTGGACAGATTAATTTCTCTCTCTCTTC 920
217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
|||||TGGAAATAATTTGCTACTATATGAATTGGACTATATGCTTCTACTTTTTC 970
234 lyIleValGlyIleIleThrPheIleLeuTrpIleCysLeuValSerAsp 250
|||||GTACTATTGGAATATTTGTTCTTTTGTGGATCTGGTTAGTAGTGAC 1020
971
```

251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLeu 267
1021 ACACACACAAACACAGAGAAATTTCCCATTTATGAAGAAGAAATACATTTCT 1070
267 uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284
1071 TTCATCATTAGAATACAGCTTCTTCACAGAAAGTCAGTCCCGTGGGTAC 1120
284 roMetLeuLysSerLeuProLeuTrpAlaIleValAlaHisPheSer 300
1121 CCATTTTAAATCCCTGCCACCTTGGCTATCGTAGTTGCACACTTTTCT 1170
301 TyrAsnTrpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetly 317
1171 TACAACCTGGACTTTTATACATTTATTTGACATTTATTCCTTACTATATGAA 1220
317 sGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValP 334
1221 GGAGATCCTTAAGGTTCAATGTTTCAGAGAAATGGGTTTATCTTCATTCG 1270
334 roTyrLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 350
1271 CTTATTAGGCTCTGGTTATGATGATCTCTCTGGTCAAGCTGCTGAC 1320
351 AsnLeuArgAlaArgTrpAsnPheSerThrLeuTrpValArgArgValPh 367
1321 AATTTAAGGGCAAAATGGAATTTTCAACTTTATGTGTCGAGAATTTT 1370
367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
1371 TAGCCTTATAGGAATGATTGGACCTGCGAGTATTCCTGGTAGCTGGCT 1420
384 heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
1421 TCATTGGCTGATATATCTTTGGCGGTGCTTCTTAACATATACACA 1470
401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspI 417
1471 ACACCTGGGAGGCTTTTGCTCTCTGATTTAGCATCAACCATCTGGATAT 1520
417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaT 434
1521 TGCTCTCGTAGCTGGTATCCTCTGGGCATCACAAATACATTTGCCA 1570
434 hrIleProGlyMetIleGlyProIleAlaArgSerLeuThrProGlu 450
1571 CTATTCAGGAATGTTGGCCCGTCATTGCTAAAGTCTGACCCCTGAT 1620
451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs 467
1621 AACACTGTTGGAGAATGGCAACCGGTCTATATTGCTGCTGCTATTAA 1670
467 nValPheGlyAlaIlePhePheThrIleuPheAlaLysGlyGluValGlnA 484
1671 TGTTTTGGTGCCATTTCTTTTACACTATTCGCCAAAGGTGAAGTACAAA 1720
484 snTrpAlaIleSerAspHisGlnGlyHisArgAsn 495
1721 ACTGGGCTCTCAATGATCACTGATGACACAGACAC 1755

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAZ50875
seq_documentation_block:
ID AAZ50875 standard; cDNA; 2602 BP.
XX
AC AAZ50875;
XX
DT 31-MAY-2000 (first entry)
XX
DE Partial human GBS toxin receptor (HP55) cDNA.
XX
KW Human GBS toxin receptor; group B beta-haemolytic streptococci; HP55;
KW pathological vascularisation; cancer metastases; angiogenesis;

KW neovascularisation; reperfusion injury; scarring; keloid;
KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW endothelial cell proliferation; antibacterial; anticancer;
KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.
XX Homo sapiens.
XX Key Location/Qualifiers
FH CDS 58..1545
FT /*tag= a
FT /product= "Partial human GBS toxin receptor"
XX
FN WO200005375-A1.
XX 03-FEB-2000.
XX 22-JUL-1999; 99WO-US16676.
XX 22-JUL-1998; 98US-0093843.
XX (UYVA-) UNIV VANDERBILT.
XX Hellerqvist CG, Fu C;
XX WPI; 2000-205377/18.
XX P-PSDB; AAY45087.
XX New polynucleotide encoding mammalian receptor for streptococcus toxin,
XX useful for diagnosis and treatment of, e.g. pneumonia in neonates -
XX
XX Claim 3; Page 77-80; 109pp; English.
XX The present cDNA sequence encodes partial human GBS (group B beta
XX -haemolytic streptococci) toxin receptor (HP55). This sequence was cloned
XX by using human embryo lung cDNA library as template. Expression vectors
XX comprising this cDNA can be transformed into host cells to express GBS
XX toxin receptor and its fragments. Detecting the receptor in tissues is
XX used to diagnose pathological vascularisation, e.g. for detecting cancer
XX metastases. GBS toxin receptors are useful for treating conditions
XX associated with pathological angiogenesis or neovascularisation
XX (specifically cancer, reperfusion injury, scarring during wound healing,
XX keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
XX neural injury), and to raise specific antibodies used for treating early
XX onset disease. Inhibitors of this receptor are useful for treating
XX pathological or hypoxia-induced endothelial cell proliferation and
XX migration.
XX
XX SQ Sequence 2602 BP; 672 A; 576 C; 583 G; 771 T; 0 other;

alignment_scores:
Quality: 2329.00 Length: 495
Ratio: 4.903 Gaps: 0
Percent Similarity: 95.960 Percent Identity: 86.263

alignment_block:
US-09-776-865-4 x AAZ50875 ..

Align seg 1/1 to: AAZ50875 from: 1 to: 2602

1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGlySe 17
58 ATGAGGTCTCCGGTTTCGAGACCTGGCCCGGAACGATGGCGAGGAGACAC 107
17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProV 34
108 GGACCGCACGCGCTTCTCTACCGGGCGCCCGCCGCGGAAGCGCTCCAG 157
34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPhe 50
158 TGTGCTGCTCTGCTCTGCTTACACTTAGCAATTTTGGCCTTTTGGTTTC 207
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67

XX New polynucleotide encoding mammalian receptor for streptococcus toxin,
 PT useful for diagnosis and treatment of, e.g. pneumonia in neonates -
 XX
 PS Claim 3; Page 89-93; 109pp; English.
 XX
 CC The present cDNA sequence encodes full length human GBS (group B beta
 CC -haemolytic streptococci) toxin receptor (HP59). This sequence was cloned
 CC by using human embryo lung cDNA library as template. Expression vectors
 CC comprising this cDNA can be transformed into host cells to express GBS
 CC toxin receptor and its fragments. Detecting the receptor in tissues is
 CC used to diagnose pathological vascularisation, e.g. for detecting cancer
 CC metastases. GBS toxin receptors are useful for treating conditions
 CC associated with pathological angiogenesis or neovascularisation
 CC (specifically cancer, reperfusion injury, scarring during wound healing,
 CC keloids, chronic inflammation (rheumatoid arthritis or psoriasis) or
 CC neural injury), and to raise specific antibodies used for treating early
 CC onset disease. Inhibitors of this receptor are useful for treating
 CC pathological or hypoxia-induced endothelial cell proliferation and
 CC migration.
 XX
 SQ Sequence 2930 BP; 729 A; 698 C; 681 G; 822 T; 0 other;

alignment_scores:
 Quality: 2329.00 Length: 495
 Ratio: 4.903 Gaps: 0
 Percent Similarity: 95.960 Percent Identity: 86.263

alignment_block:

US-09-776-865-4 x AAZ50879 ..

Align seg 1/1 to: AAZ50879 from: 1 to: 2930

1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGlyLeu 17
 386 ATGAGGTCTCCGGTTCGAGACCTGGCCCGGACGATGGCGAGGAGAGAC 435
 17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProv 34
 436 GGACCCGACGCGCTTCTACCGGGCGCCCGCCACGGCGCGAAGCGCTCCAG 485
 34 aLcysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPhe 50
 486 TGTGCTGCTCTGCTCGTTCAACTTAGCAATTTGGCCCTTTTGGCTTC 535
 51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspHe 67
 536 TTCATTGTGTATGCTACGTGGAATCTGAGTGTGGTTCGCTAGTGATAT 585
 67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
 586 GGTAGATTCAAAATACAACTTTAGAACATAATAGAACTTCAAGCGCGTGC 635
 84 laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
 636 CAGAGCATTTCTGCTCCCATAAAGTTTCATCATATAATCAAAACGGGTAAGAAG 685
 101 TyrArgTyrAspAlaGluThrGlnGlyTyrPheLeuGlySerPhePheTy 117
 686 TACCAATGGGATGCGAAGAACTCAAGGATGGATTTCTCGGTTCCTTTTATA 735
 117 rGlyTyrIleLeuThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134
 736 TGGCTACATCATCACACAGATTCCTGGAGGATATGTTGCCAGCAAAATAG 785
 134 lyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
 786 GGGGGAATAATGCTGTAGGATTTGGGATCTTGGCACTGCTGCTCACC 835
 151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
 836 CTGTTCACTCCCATTCGTCAGATTTAGGAGTTGGACCACCTCATTTGTA 885

167 uArgAlaLeuGluGlyLeuGlyGluGlyValThrTyrProAlaMethHisA 184
 886 CAGACGACTAGAAAGGACTAGGAGGGGTGTACATTTCCAGCCATGCATG 935
 184 laMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
 936 CCATGTGGTCTCTTGGGCTCCCTTGAAGAAGCAAACTCTTAGC 985
 201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
 986 ATTTCGTATGCAGAGCACAGCTTGGACAGTAATTTCTCTCTCTTC 1035
 217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
 1036 TGAATAAATTTGCTACTATATGAATGGACTTATGTCTTCTACTTTT 1085
 234 lyIleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
 1086 GTACTATTGGAATATTTTGGTTCTTTTGGATCTGGTTAGTTAGTGAC 1135
 251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
 1136 ACACACAAAAACACAGAGAAATTTCCCATATGAAAGGAATACATTCT 1185
 267 uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284
 1186 TTCATCATTAAGAAATCAGCTTTCTTCACAGAAATCAGTCCGCGGTG 1235
 284 roMetLeuLysSerLeuProLeuTrpAlaIleValValAlaHisPheSer 300
 1236 CCATTTAAATCCCTGCCACTTTGGCTATCGTAGTTGCACACTTTTCT 1285
 301 TyrAsnTrpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLy 317
 1286 TCAACTGGACTTTTATACCTTATTATGACATTAATGCCTACTTATATGA 1335
 317 sGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValP 334
 1336 GGAGATCTTAAGTTCATGTTCAAGAGAAATGGTTTATCTTCATTGTC 1385
 334 roTyrLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 350
 1386 CTTATTAGGCTCTCTGTTATGATGATCTGCTGCTCAAGCTGCTGAC 1435
 351 AsnLeuArgAlaArgTrpAsnPheSerThrLeuTrpValArgArgValPh 367
 1436 AATTTAAGGCAAAATGGAATTTTCACTTTATGTGTTCGCAGAAATTT 1485
 367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
 1486 TAGCCTTATAGGAATGATTGGACCTGCAGTATTCTCTGGTAGCTGGCT 1535
 384 heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
 1536 TCATTGGCTGTGATTATTTCTTGGCGGTTGCTTTCTTAAGTATATCA 1585
 401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspI 417
 1586 ACCTGGAGGCTTTTGTCTTCTGGATTAGCATCAACCATCTGGATAT 1635
 417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAla 434
 1636 TGCTCTCGTATGCTGGTATCTCTGGCATCACAAATACATTGGCCA 1685
 434 hrIleProGlyMetIleGlyProIleLeuAlaArgSerLeuThrProGlu 450
 1686 CTATTCAGGAATGGTGGCGCGCTCATTTGCTAAAAAGTCTGACCCCTG 1735
 451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs 467
 1736 AACACTGTTGGAGAAATGGCAACCGGTTCCTATATTGCTGCTGCTATTAA 1785

467 nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnA 484
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 1835
 1786 TGTGGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 1835
 484 snTrpAlaIleSerAspHisGlnGlyHisArgAsn 495
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 1835
 1836 ACTGGGCTCTCAATGATCAGCATGGACAGACAC 1870

seq_name: /SDS1/gcgdata/hold-geneseq/geneseq-n-emb1/NA2001A.DAT: AAD10325

seq_documentation_block:

ID AAD10325 standard; DNA; 2930 BP.

AC AAD10325;

DT 16-OCT-2001 (first entry)

DE Human group B beta-haemolytic Streptococci toxin receptor (HP59) DNA.

KW Human; group B beta-haemolytic Streptococci toxin receptor; GBS; HP59;
 KW cytostatic; vulnary; antiatherosclerotic; osteopathic; vasotropic;
 KW prevention; attenuation; pathoangiogenic condition; cancer; scar;
 KW wound healing; gliosis; nerve injury; chronic wound; reperfusion injury;
 KW keloid; rheumatoid arthritis; atherosclerosis; osteoarthritis; psoriasis;
 KW vaccine; ds.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS

FT 263..1873

FT /*tag= a

FT /product= "Human GBS toxin receptor protein (HP59) "

PN WO200156598-A2.

XX 09-AUG-2001.

XX 02-FEB-2001; 2001WO-US03662.

XX 02-FEB-2000; 2000US-0179870.

XX (UYVA-) UNIV VANDERBILT.

XX Hellerqvist CG;

XX WPI: 2001-488844/53.

XX P-PSDB; AAE06518.

PT Preventing or attenuating pathoangiogenic conditions e.g. cancer,
 PT chronic wounds, osteoarthritis, keloids and psoriasis in a mammal, by
 PT administering group B beta-hemolytic Streptococci toxin receptor or its
 PT fragment -

PS Disclosure: Page 41-44; 52pp; English.

XX The present sequence is a DNA encoding full length human group B beta-
 CC haemolytic Streptococci (GBS) toxin receptor protein, HP59. The present
 CC invention relates to a method for preventing or attenuating a patho-
 CC angiogenic condition in a mammal which comprises administering a patho-
 CC to induce or maintain an immune response to one of GBS toxin receptors.
 CC The method is useful for preventing or ameliorating pathoangiogenic
 CC conditions such as cancer, scarring during wound healing, gliosis
 CC during repair of nerve injury, chronic wounds, keloids, reperfusion
 CC injury, rheumatoid arthritis, atherosclerosis, osteoarthritis and
 CC psoriasis in mammals. The proteins of the invention are also used
 CC as vaccines.

SQ Sequence 2930 BP; 729 A; 698 C; 681 G; 822 T; 0 other;

alignment_scores:

Quality: 2329.00

Length: 495

Ratio: 4.903 Gaps: 0
 Percent Similarity: 95.960 Percent Identity: 86.263
 alignment_block:
 US-09-776-865-4 x AAD10325 ..
 Align seg 1/1 to: AAD10325 from: 1 to: 2930

1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGlySe 17
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 435
 386 ATGAGGTCTCCGGTTCGAGACCTGGCCCGGACGATGGCGGAGGAGAC 435
 17 rAspArgThrProLeuGlnArgAlaProArgAlaGluProAlaPro 34
 :|||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 485
 436 GGACCGCAGCCCTCTTCTACCGGGCGCCACGCGGCGAAGCGCTCCAG 485
 34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPhe 50
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 535
 486 TGTGCTGCTCTGCTCGTTCACAACTTAGCAATTTTGGCCTTTTGGTTTC 535
 51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAsp 67
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 585
 536 TTCATTGTGTATGATACGTGTAATCTGAGTGTGCGTGTAGTGATAT 585
 67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCys 84
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 635
 586 GGTAGATTCAATACAACTTTAGAGATAATAGAACTTCCAAGCGGTGC 635
 84 laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 685
 636 CAGAGCATTCGTGCTCCATATAAGTTTCATCATATAAAGCGGTAAAG 685
 101 TyrArgTrpAspAlaGluThrGlnGlyTyrIleLeuGlySerPhePhe 117
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 735
 686 TACCAATGGGATGCAGAACTCAAGGATGGATTCTCGGTTCTCTTTTAA 735
 117 rGlyTyrIleThrGlnIleProGlyGlyTyrValAlaLaserArgSerG 134
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 785
 736 TGGCTACATCATCACAGATTCTTGAGGATATGTTGCCAGCAAAATAG 785
 134 lyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 835
 786 GGGGGAATAATGCTGCTAGGATTGGGATCCTTGGCACCTGCTGCTCACC 835
 151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAla 167
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 885
 836 CTGTTTCACTCCCATTCGTCAGATTAGGAGTTGGACCATCTCATTGTACT 885
 167 uArgAlaLeuGlyGlyLeuGlyGlyValThrTyrProAlaMethHisA 184
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 935
 886 CAGAGCAGTAGAAGGACTAGGAGAGGTGTACATTTCAGCCATGCATG 935
 184 laMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 985
 936 CCATGTGGTCTTCTTGGGCTCCCTCTTGAAGAAGCAAACTTCTTAGC 985
 201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeu 217
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 1035
 986 ATTTCGATGTCAGGACGACAGCTTGGGACAGTAATTTCTCTCTCTTTC 1035
 217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePhe 234
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 1085
 1036 TGGAAATAATTTGCTACTATATGATGAGTTGGACTTATGCTCTTACTTTTTC 1085
 234 lyTleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 1135
 1086 GTACTATTGGAATATTTGGTTCTTTTGGGATCTGTTAGTTAGTGAC 1135
 251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
 |||||TGTGGTGGCCATTCTTTACACTATTTCGCAAGAGTGAAGTACAAA 1185
 1136 ACACCAACAAAAACAGAGAAATTTCCCATTTATGAAAGGAATACATTCT 1185
 267 uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284


```
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
151 TTCAATGTGTATGCATTACGTTCAATCTGAATCTGAGTCTGCGTTAGTGGATAT 200
67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
201 GGTAGATTCAAAATCAACTTTAGAAGATAATAGAACTTCCAAGCGGTGC 250
84 LaGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
251 CAGAGCATTCCTGCCCAATAAAGTTTCATCATATAATCAACGGGTAAAG 300
101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPheTyr 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
301 TACCAATGGGATGCAGAAACTCAAGGATGGATTCTCGSTTCCITTTTAA 350
117 rGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSerg 134
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
351 TGGCTACATCATCACAGAGATTCTCGGAGGATATGTTGCCAGCAAAATAG 400
134 LyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
401 GGGGAAAAATGCTCTAGGATTTGGGATTCCTGGCACTGCTGCTCCTCACC 450
151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
451 CTGTTCACTCCCATTTGTCGACAGATTAGGAGTTGGACCACTCATTTGACT 500
167 uArgAlaLeuGluGlyLeuGlyGluGlyValThrTyrProAlaMethHisA 184
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
501 CAGAGCACTAGAAAGACTAGGAGAGGTTGTACATTTCCAGCCATGCATG 550
184 laMetTrpSerSerTrpAlaProLeuGluArgSerLysLeuLeuSer 200
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
551 CCATGTGCTCTCTTGGGCTCCCCCTCTTGAAGAAGCAAACTTCCTTAGC 600
201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
601 ATTTTCATATGCAGGAGCACAGCTTGGGACAGTAATTTCTCTCTCTCTTC 650
217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
651 TGGAAATAATTGCTACATATATGAATTGGACTTATGCTTCTACTCTTTTGT 700
234 LyIleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
701 GTACTATTGGAAATATTTGGTTCTTTTGGATCTGCTAGTTAGTGAC 750
251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
751 ACACCACAAAAACACAGAGAAATTTCCCATTTATGAAAAGGAATACATCT 800
267 uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
801 TTTCATCATTAAGAAATCAGCTTTCTTCACAGAGTCAGGCCGCTGGGTAC 850
284 roMetLeuLysSerLeuProLeuTrpAlaIleValValAlaHisPheSer 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
851 CCATTTTAAATCCCTGCCACTTTGGGCTATCGTAGTTGCACACTTTTCT 900
301 TyrAsnTrpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLy 317
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
901 TACAACCTGGACTTTTATATCTTTTATTCACATTAATTGGCTACTTATATGAA 950
317 sGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValP 334
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
951 GGAGATCCTAAGGTCAATGCTCAAGAGAATGGGTTTTTATCTTCATTGC 1000
334 roTyrLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAsp 350
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1001 CTATTTAGGCTCTTGTTATGATGATCTGCTGCTGCTCAAGCTGCTGAC 1050
351 AsnLeuArgAlaArgTrpAsnPheSerThrLeuTrpValArgArgValPh 367
```

```
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1051 AATTTAAGGGCAAAATGGAATTTTCAACTTTATGTCTGCCAGAAATTTT 1100
367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1101 TAGCCTTTATAGGAATGATTGGACCTGCAGTATTCCCTGCTAGCTGCTGGCT 1150
384 heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1151 TCATTGGCTGTGATTATCTTTGGCCGTTGCTTTCTTAACATATATCAACA 1200
401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIl 417
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1201 AACTGGGAGGCTTTTGTCTCTCTGGATTAGCATCAACCATCTGGATAT 1250
417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaT 434
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1251 TGTCTCTTCGTATGCTGTTATCTCTCTCCGGGCATCACAAATACATTTGCCA 1300
434 hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1301 CTATTCAGGAATGGTTGGGCCGCTCATTTGCTAAAGTCTGACCCCTGAT 1350
451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs 467
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1351 AACACTGTTGGAGAAATGGCAACCGTGTCTATATTGCTGCTGCTATTAA 1400
467 nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnA 484
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1401 TGTTTTTGGTGCCATTTTCTTTACACTATTTCGCCAAAGGTGAAGTACAAA 1450
484 snTrpAlaIleSerAspHisGlnGlyHisArgAsn 495
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1451 ACTGGGCTCTCATGATCACCATGGACACAGACAC 1485
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-n-emb1/NA2001A.DAT:AAK94876
seq_documentation_block:
ID AAK94876 standard; cDNA; 2712 BP.
XX
AC AAK94876;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human full-length cDNA, SEQ ID NO: 4068.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
XX
PR 11-JAN-2000; 2000JP-0118774.
XX
PR 02-MAY-2000; 2000JP-0183765.
XX
(HELI-) HELIX RES INST.
PA
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.
DR P-PSDB; AAM93914.
XX
PT 830 primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -
PS Claim 8; SEQ ID NO 4068; 1380pp + sequence listing; English.
XX
```



```
467 AsnValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGI 483
|||||
1583 AATGTTTTGGTCACATTTCTTTACACTATTGCGCAAGGTGAAGTACA 1632
|||||
483 nAsnTrpAlaIleSerAspHisGlnGlyHisArgAsn 495
|||||
1633 AAACGGGCTCTCATCATGATCACCATTGACACAGACAC 1669

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ50880
seq_documentation_block:
ID  AAZ50880 standard; cDNA; 1485 BP.
XX
AC  AAZ50880;
XX
DT  31-MAY-2000 (first entry)
XX
DE  Human/Sheep consensus GBS toxin receptor cDNA sequence-1.
XX
KW  Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59;
KW  pathological vascularisation; cancer metastases; angiogenesis;
KW  neovascularisation; reperfusion injury; scarring; keloid;
KW  chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;
KW  endothelial cell proliferation; antibacterial; anticancer;
KW  anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.
XX
OS  Homo sapiens.
OS  Ovis sp.
XX
FH  Key
FT  mat_peptide
FT  1..1485
FT  Location/Qualifiers
FT  /tag= a
FT  /product= "Human/sheep consensus GBS toxin receptor
FT  protein sequence"
FT  /transl_except= (pos:4..6; aa:Xaa)
FT  /transl_except= (pos:7..9; aa:Xaa)
FT  /transl_except= (pos:16..18; aa:Xaa)
FT  /transl_except= (pos:22..24; aa:Xaa)
FT  /transl_except= (pos:28..30; aa:Xaa)
FT  /transl_except= (pos:31..33; aa:Xaa)
FT  /transl_except= (pos:34..36; aa:Xaa)
FT  /transl_except= (pos:46..48; aa:Xaa)
FT  /transl_except= (pos:49..51; aa:Xaa)
FT  /transl_except= (pos:58..60; aa:Xaa)
FT  /transl_except= (pos:61..63; aa:Xaa)
FT  /transl_except= (pos:64..66; aa:Xaa)
FT  /transl_except= (pos:67..69; aa:Xaa)
FT  /transl_except= (pos:70..72; aa:Xaa)
FT  /transl_except= (pos:73..75; aa:Xaa)
FT  /transl_except= (pos:79..81; aa:Xaa)
FT  /transl_except= (pos:85..87; aa:Xaa)
FT  /transl_except= (pos:91..93; aa:Xaa)
FT  /transl_except= (pos:100..102; aa:Xaa)
FT  /transl_except= (pos:124..126; aa:Xaa)
FT  /transl_except= (pos:130..132; aa:Xaa)
FT  /transl_except= (pos:136..138; aa:Xaa)
FT  /transl_except= (pos:154..156; aa:Xaa)
FT  /transl_except= (pos:157..159; aa:Xaa)
FT  /transl_except= (pos:163..165; aa:Xaa)
FT  /transl_except= (pos:169..171; aa:Xaa)
FT  /transl_except= (pos:181..183; aa:Xaa)
FT  /transl_except= (pos:187..189; aa:Xaa)
FT  /transl_except= (pos:190..192; aa:Xaa)
FT  /transl_except= (pos:196..198; aa:Xaa)
FT  /transl_except= (pos:202..204; aa:Xaa)
FT  /transl_except= (pos:211..213; aa:Xaa)
FT  /transl_except= (pos:220..222; aa:Xaa)
FT  /transl_except= (pos:223..225; aa:Xaa)
FT  /transl_except= (pos:235..237; aa:Xaa)
FT  /transl_except= (pos:241..243; aa:Xaa)
FT  /transl_except= (pos:244..246; aa:Xaa)
FT  /transl_except= (pos:250..252; aa:Xaa)
FT  /transl_except= (pos:277..279; aa:Xaa)
FT  /transl_except= (pos:280..282; aa:Xaa)
FT  /transl_except= (pos:283..285; aa:Xaa)
FT  /transl_except= (pos:295..297; aa:Xaa)
FT  /transl_except= (pos:304..306; aa:Xaa)
FT  /transl_except= (pos:337..339; aa:Xaa)
FT  /transl_except= (pos:340..342; aa:Xaa)
FT  /transl_except= (pos:346..348; aa:Xaa)
FT  /transl_except= (pos:367..369; aa:Xaa)
FT  /transl_except= (pos:394..396; aa:Xaa)
FT  /transl_except= (pos:397..399; aa:Xaa)
FT  /transl_except= (pos:406..408; aa:Xaa)
FT  /transl_except= (pos:409..411; aa:Xaa)
FT  /transl_except= (pos:412..414; aa:Xaa)
FT  /transl_except= (pos:421..423; aa:Xaa)
FT  /transl_except= (pos:430..432; aa:Xaa)
FT  /transl_except= (pos:433..435; aa:Xaa)
FT  /transl_except= (pos:436..438; aa:Xaa)
FT  /transl_except= (pos:442..444; aa:Xaa)
FT  /transl_except= (pos:445..447; aa:Xaa)
FT  /transl_except= (pos:463..465; aa:Xaa)
FT  /transl_except= (pos:475..477; aa:Xaa)
FT  /transl_except= (pos:481..483; aa:Xaa)
FT  /transl_except= (pos:487..489; aa:Xaa)
FT  /transl_except= (pos:490..492; aa:Xaa)
FT  /transl_except= (pos:493..495; aa:Xaa)
FT  /transl_except= (pos:496..498; aa:Xaa)
FT  /transl_except= (pos:502..504; aa:Xaa)
FT  /transl_except= (pos:514..516; aa:Xaa)
FT  /transl_except= (pos:529..531; aa:Xaa)
FT  /transl_except= (pos:535..537; aa:Xaa)
FT  /transl_except= (pos:562..564; aa:Xaa)
FT  /transl_except= (pos:589..591; aa:Xaa)
FT  /transl_except= (pos:595..597; aa:Xaa)
FT  /transl_except= (pos:598..600; aa:Xaa)
FT  /transl_except= (pos:604..606; aa:Xaa)
FT  /transl_except= (pos:619..621; aa:Xaa)
FT  /transl_except= (pos:634..636; aa:Xaa)
FT  /transl_except= (pos:655..657; aa:Xaa)
FT  /transl_except= (pos:691..693; aa:Xaa)
FT  /transl_except= (pos:694..696; aa:Xaa)
FT  /transl_except= (pos:700..702; aa:Xaa)
FT  /transl_except= (pos:703..705; aa:Xaa)
FT  /transl_except= (pos:706..708; aa:Xaa)
FT  /transl_except= (pos:712..714; aa:Xaa)
FT  /transl_except= (pos:715..717; aa:Xaa)
FT  /transl_except= (pos:724..726; aa:Xaa)
FT  /transl_except= (pos:727..729; aa:Xaa)
FT  /transl_except= (pos:736..738; aa:Xaa)
FT  /transl_except= (pos:748..750; aa:Xaa)
FT  /transl_except= (pos:757..759; aa:Xaa)
FT  /transl_except= (pos:760..762; aa:Xaa)
FT  /transl_except= (pos:769..771; aa:Xaa)
FT  /transl_except= (pos:772..774; aa:Xaa)
FT  /transl_except= (pos:775..777; aa:Xaa)
FT  /transl_except= (pos:778..780; aa:Xaa)
FT  /transl_except= (pos:790..792; aa:Xaa)
FT  /transl_except= (pos:793..795; aa:Xaa)
FT  /transl_except= (pos:811..813; aa:Xaa)
FT  /transl_except= (pos:820..822; aa:Xaa)
FT  /transl_except= (pos:847..849; aa:Xaa)
FT  /transl_except= (pos:850..852; aa:Xaa)
FT  /transl_except= (pos:853..855; aa:Xaa)
FT  /transl_except= (pos:856..858; aa:Xaa)
FT  /transl_except= (pos:862..864; aa:Xaa)
FT  /transl_except= (pos:880..882; aa:Xaa)
FT  /transl_except= (pos:883..885; aa:Xaa)
FT  /transl_except= (pos:892..894; aa:Xaa)
FT  /transl_except= (pos:922..924; aa:Xaa)
FT  /transl_except= (pos:928..930; aa:Xaa)
FT  /transl_except= (pos:943..945; aa:Xaa)
FT  /transl_except= (pos:952..954; aa:Xaa)
FT  /transl_except= (pos:955..957; aa:Xaa)
FT  /transl_except= (pos:970..972; aa:Xaa)
```



```
851 CYATKYTRAAATCMCTGCCACTTTGGGCTATYGTMTTGCACAYTTTTCT 900
301 TyAsnTprThrPheTyThrLeuLeuThrLeuLeuProThrTyMetLy 317
|||||
901 TACAACTGGACTTTTATACATTTTGTGACMTATTTGCTACTTAYATGAA 950
|||||
317 sGluValLeuArgPheAsnIleGlnGlnAsnGlyPheLeuSerAlaValP 334
|||||
951 GGARRTCCTCAAGTTCATATTCATCAAGACAATGGGTTTTTATCTCAKATSC 1000
|||||
334 roTyLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaasp 350
|||||
1001 CTTATTTAGGTTSTTGGTTATGTATGATCCTGCTCKGGTCAAGCTGCTGAC 1050
|||||
351 AsnLeuArgAlaArgTprAsnPheSerThrLeuTrpValArgArgValPh 367
|||||
1051 AATTAAAGGGCAARATGGAATTTTCAACTTTRTGKGTTCGMAGARTTTT 1100
|||||
367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
|||||
1101 TAGCCTTATAGRATGATTGGACCTGCRRTATTCCTGTGTCGCGWGMT 1150
|||||
384 heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
|||||
1151 TYATWGGCTGTGATATTCYTTGGCYGTTCGTTCTCTTAACVATATCAACA 1200
|||||
401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIl 417
|||||
1201 ACMCTGGAGGCTTTTGCTCTCTGGATTTAGCATCAACCATCTGGAYAT 1250
|||||
417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlat 434
|||||
1251 TGCTCTCTGATGCTGTATYCTCTGGGCATCACAAATACMTTGGCA 1300
|||||
434 hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
|||||
1301 CTATTCCKGSAATGRTTGGGCCCTCATCTGCTYARAAGTCTKACCCCTGAK 1350
|||||
451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaAlaIleAs 467
|||||
1351 AACACTRTGGAGATGGCAACVGTGTTCTRYATYGTGCTGCTGTATYAA 1400
|||||
467 nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnA 484
|||||
1401 TGTWTTTGGTGCCATTTCTTYACACTATTGCGCAAGGTCAAGTRCAAA 1450
|||||
484 snTrpAlaIleSerAspHisGlnGlyHisArg 494
|||||
1451 ACTGGCYMTCAATCATCACCAGGACACAGA 1482
|||||
```

seq_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.AAZ50881

seq_documentation_block:

ID AAZ50881 standard; cDNA; 1485 BP.

XX AC AAZ50881;

XX DT 31-MAY-2000 (first entry)

XX DE Human/Sheep consensus GBS toxin receptor cDNA sequence-2.

XX KW Human GBS toxin receptor; group B beta-haemolytic streptococci; HP59;

XX KW pathological vascularisation; cancer metastases; angiogenesis;

XX KW neovascularisation; reperfusion injury; scarring; keloid;

XX KW chronic inflammation; rheumatoid arthritis; psoriasis; neural injury;

XX KW endothelial cell proliferation; antibacterial; anticancer;

XX KW anti-angiogenic; anti-inflammatory; anti-arthritis; anti-psoriatic; ss.

XX OS Homo sapiens.

XX OS Ovis sp.

XX FH Key

XX mat_peptide 1..1485

XX Location/Qualifiers

FT

```
/*tag= a
/product= "Human/sheep consensus GBS toxin receptor
protein sequence"
/transl_except= (pos:4..6; aa:Xaa)
/transl_except= (pos:7..9; aa:Xaa)
/transl_except= (pos:16..18; aa:Xaa)
/transl_except= (pos:22..24; aa:Xaa)
/transl_except= (pos:28..30; aa:Xaa)
/transl_except= (pos:31..33; aa:Xaa)
/transl_except= (pos:34..36; aa:Xaa)
/transl_except= (pos:46..48; aa:Xaa)
/transl_except= (pos:49..51; aa:Xaa)
/transl_except= (pos:58..60; aa:Xaa)
/transl_except= (pos:61..63; aa:Xaa)
/transl_except= (pos:64..66; aa:Xaa)
/transl_except= (pos:67..69; aa:Xaa)
/transl_except= (pos:70..72; aa:Xaa)
/transl_except= (pos:73..75; aa:Xaa)
/transl_except= (pos:79..81; aa:Xaa)
/transl_except= (pos:85..87; aa:Xaa)
/transl_except= (pos:91..93; aa:Xaa)
/transl_except= (pos:100..102; aa:Xaa)
/transl_except= (pos:124..126; aa:Xaa)
/transl_except= (pos:130..132; aa:Xaa)
/transl_except= (pos:136..138; aa:Xaa)
/transl_except= (pos:154..156; aa:Xaa)
/transl_except= (pos:157..159; aa:Xaa)
/transl_except= (pos:163..165; aa:Xaa)
/transl_except= (pos:169..171; aa:Xaa)
/transl_except= (pos:181..183; aa:Xaa)
/transl_except= (pos:187..189; aa:Xaa)
/transl_except= (pos:190..192; aa:Xaa)
/transl_except= (pos:196..198; aa:Xaa)
/transl_except= (pos:202..204; aa:Xaa)
/transl_except= (pos:211..213; aa:Xaa)
/transl_except= (pos:220..222; aa:Xaa)
/transl_except= (pos:223..225; aa:Xaa)
/transl_except= (pos:235..237; aa:Xaa)
/transl_except= (pos:241..243; aa:Xaa)
/transl_except= (pos:244..246; aa:Xaa)
/transl_except= (pos:250..252; aa:Xaa)
/transl_except= (pos:277..279; aa:Xaa)
/transl_except= (pos:280..282; aa:Xaa)
/transl_except= (pos:283..285; aa:Xaa)
/transl_except= (pos:295..297; aa:Xaa)
/transl_except= (pos:304..306; aa:Xaa)
/transl_except= (pos:337..339; aa:Xaa)
/transl_except= (pos:340..342; aa:Xaa)
/transl_except= (pos:346..348; aa:Xaa)
/transl_except= (pos:367..369; aa:Xaa)
/transl_except= (pos:394..396; aa:Xaa)
/transl_except= (pos:397..399; aa:Xaa)
/transl_except= (pos:406..408; aa:Xaa)
/transl_except= (pos:409..411; aa:Xaa)
/transl_except= (pos:412..414; aa:Xaa)
/transl_except= (pos:421..423; aa:Xaa)
/transl_except= (pos:430..432; aa:Xaa)
/transl_except= (pos:433..435; aa:Xaa)
/transl_except= (pos:442..444; aa:Xaa)
/transl_except= (pos:445..447; aa:Xaa)
/transl_except= (pos:463..465; aa:Xaa)
/transl_except= (pos:475..477; aa:Xaa)
/transl_except= (pos:481..483; aa:Xaa)
/transl_except= (pos:487..489; aa:Xaa)
/transl_except= (pos:490..492; aa:Xaa)
/transl_except= (pos:493..495; aa:Xaa)
/transl_except= (pos:496..498; aa:Xaa)
/transl_except= (pos:502..504; aa:Xaa)
/transl_except= (pos:514..516; aa:Xaa)
/transl_except= (pos:529..531; aa:Xaa)
/transl_except= (pos:535..537; aa:Xaa)
/transl_except= (pos:562..564; aa:Xaa)
```



```
FT /transl_except= (pos:589..591; aa:Xaa)
FT /transl_except= (pos:595..597; aa:Xaa)
FT /transl_except= (pos:598..600; aa:Xaa)
FT /transl_except= (pos:604..606; aa:Xaa)
FT /transl_except= (pos:619..621; aa:Xaa)
FT /transl_except= (pos:634..636; aa:Xaa)
FT /transl_except= (pos:655..657; aa:Xaa)
FT /transl_except= (pos:691..693; aa:Xaa)
FT /transl_except= (pos:694..696; aa:Xaa)
FT /transl_except= (pos:700..702; aa:Xaa)
FT /transl_except= (pos:703..705; aa:Xaa)
FT /transl_except= (pos:706..708; aa:Xaa)
FT /transl_except= (pos:712..714; aa:Xaa)
FT /transl_except= (pos:715..717; aa:Xaa)
FT /transl_except= (pos:724..726; aa:Xaa)
FT /transl_except= (pos:727..729; aa:Xaa)
FT /transl_except= (pos:736..738; aa:Xaa)
FT /transl_except= (pos:748..750; aa:Xaa)
FT /transl_except= (pos:757..759; aa:Xaa)
FT /transl_except= (pos:760..762; aa:Xaa)
FT /transl_except= (pos:769..771; aa:Xaa)
FT /transl_except= (pos:772..774; aa:Xaa)
FT /transl_except= (pos:775..777; aa:Xaa)
FT /transl_except= (pos:778..780; aa:Xaa)
FT /transl_except= (pos:790..792; aa:Xaa)
FT /transl_except= (pos:793..795; aa:Xaa)
FT /transl_except= (pos:811..813; aa:Xaa)
FT /transl_except= (pos:820..822; aa:Xaa)
FT /transl_except= (pos:847..849; aa:Xaa)
FT /transl_except= (pos:850..852; aa:Xaa)
FT /transl_except= (pos:853..855; aa:Xaa)
FT /transl_except= (pos:856..858; aa:Xaa)
FT /transl_except= (pos:862..864; aa:Xaa)
FT /transl_except= (pos:880..882; aa:Xaa)
FT /transl_except= (pos:883..885; aa:Xaa)
FT /transl_except= (pos:892..894; aa:Xaa)
FT /transl_except= (pos:922..924; aa:Xaa)
FT /transl_except= (pos:928..930; aa:Xaa)
FT /transl_except= (pos:943..945; aa:Xaa)
FT /transl_except= (pos:952..954; aa:Xaa)
FT /transl_except= (pos:955..957; aa:Xaa)
FT /transl_except= (pos:970..972; aa:Xaa)
FT /transl_except= (pos:994..996; aa:Xaa)
FT /transl_except= (pos:997..999; aa:Xaa)
FT /transl_except= (pos:1009..1011; aa:Xaa)
FT /transl_except= (pos:1012..1014; aa:Xaa)
FT /transl_except= (pos:1033..1035; aa:Xaa)
FT /transl_except= (pos:1063..1065; aa:Xaa)
FT /transl_except= (pos:1081..1083; aa:Xaa)
FT /transl_except= (pos:1084..1086; aa:Xaa)
FT /transl_except= (pos:1090..1092; aa:Xaa)
FT /transl_except= (pos:1096..1098; aa:Xaa)
FT /transl_except= (pos:1111..1113; aa:Xaa)
FT /transl_except= (pos:1126..1128; aa:Xaa)
FT /transl_except= (pos:1129..1131; aa:Xaa)
FT /transl_except= (pos:1138..1140; aa:Xaa)
FT /transl_except= (pos:1141..1143; aa:Xaa)
FT /transl_except= (pos:1144..1146; aa:Xaa)
FT /transl_except= (pos:1147..1149; aa:Xaa)
FT /transl_except= (pos:1150..1152; aa:Xaa)
FT /transl_except= (pos:1153..1155; aa:Xaa)
FT /transl_except= (pos:1168..1170; aa:Xaa)
FT /transl_except= (pos:1174..1176; aa:Xaa)
FT /transl_except= (pos:1180..1182; aa:Xaa)
FT /transl_except= (pos:1189..1191; aa:Xaa)
FT /transl_except= (pos:1201..1203; aa:Xaa)
FT /transl_except= (pos:1246..1248; aa:Xaa)
FT /transl_except= (pos:1270..1272; aa:Xaa)
FT /transl_except= (pos:1291..1293; aa:Xaa)
FT /transl_except= (pos:1306..1308; aa:Xaa)
FT /transl_except= (pos:1315..1317; aa:Xaa)
FT /transl_except= (pos:1324..1326; aa:Xaa)
FT /transl_except= (pos:1330..1332; aa:Xaa)
```

```
FT /transl_except= (pos:1333..1335; aa:Xaa)
FT /transl_except= (pos:1339..1341; aa:Xaa)
FT /transl_except= (pos:1348..1350; aa:Xaa)
FT /transl_except= (pos:1357..1359; aa:Xaa)
FT /transl_except= (pos:1372..1374; aa:Xaa)
FT /transl_except= (pos:1375..1377; aa:Xaa)
FT /transl_except= (pos:1381..1383; aa:Xaa)
FT /transl_except= (pos:1384..1386; aa:Xaa)
FT /transl_except= (pos:1396..1398; aa:Xaa)
FT /transl_except= (pos:1402..1404; aa:Xaa)
FT /transl_except= (pos:1420..1422; aa:Xaa)
FT /transl_except= (pos:1444..1446; aa:Xaa)
FT /transl_except= (pos:1456..1458; aa:Xaa)
FT /transl_except= (pos:1459..1461; aa:Xaa)
FT /transl_except= (pos:1462..1464; aa:Xaa)
FT /transl_except= (pos:1471..1473; aa:Xaa)
FT /transl_except= (pos:1483..1485; aa:Xaa)
FT /note= "Xaa = unknown"
XX
PN WO200005375-A1.
XX
XX 03-FEB-2000.
PD
XX
XX 22-JUL-1999; 99WO-US16676.
PF
XX
XX 22-JUL-1998; 98US-0093843.
PR
XX
XX (UYVA-) UNIV VANDERBILT.
PA
XX
XX PI Hellerqvist CG, Fu C;
XX
XX DR WPI; 2000-205377/18.
XX
XX DR P-PSDB; AAY45090.
XX
alignment_scores:
    Quality: 1971.00      Length: 494
    Ratio: 4.879         Gaps: 0
    Percent Similarity: 81.781      Percent Identity: 78.138
alignment_block:
US-09-776-865-4 x AAZ50881 ..
Align seg 1/1 to: AAZ50881 from: 1 to: 1485
1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGlySe 17
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 ATGANGTCNCCGGTTNNNGACNTNGCCNGANGCANGCGGAGGAGNGCNC 50
   :||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProV 34
   :||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 GGACCGCACNCCNCTNCTNCNGNGCGCCCGCGGCGNGAANCCGCTCCAG 100
34 alCysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyPhe 50
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 TNGTCTGCTCTGCTGCTACACNTAGCANTTTTGNCTTTTGTGTTTC 150
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAsp 67
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 TTNCTNTNTATNCATTCAGNGTGAATCTGAGNGTTCGNNTAGTGGANAT 200
67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCys 84
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 GGTNGATTCAANACAACCTNNNAAGATAATACACNTCCNANGNGTGN 250
84 laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLys 100
   :||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 CAGAGCATTCCTGCTCCATAAAAGTTCTCANAANCAACGCGTAANAAG 300
101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePhe 117
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 TACCNTTGGGATCAGAAACTCAAGGATGATTCTCGGNTCNTTTTNTA 350
```

117 rGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134
|||||
351 TGGCTACATCATCACANATTCCTGGAGGATATGTTGCCAGCAANNG 400
134 lyGlyLysLeuLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
|||||
401 GGGGAANNNTGCTAGGATTTGGGATCTNTGNACNGCTNTCTCACC 450
151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
|||||
451 CTGTTCACTCCCTNNGCGAGCATTTGGAGTNGGANCCTNTTGNACT 500
167 uArgAlaLeuGlyLeuGlyGlyValThrTyrProAlaMethIle 184
|||||
501 CAGNGCACTAGCAAGGCTAGGAGAGGGTGTNACATNTCCAGCCATGCATG 550
184 laMetTrpSerSerTrpAlaProProLeuGluArgSerLysLeuSer 200
|||||
551 CCATGTGGTCTTCNTGGGCTCCCTCTTGAAGAAGCAANCTTCTNAGN 600
201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
|||||
601 ATTCNTATGAGAGGACACANCTTGGACAGATNTTCTCTCTCTCTTC 650
217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
|||||
651 TGGANTAATTTGCTACTATATGAATGGACTTATGCTCTCTANTTTTG 700
234 lyIleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
|||||
701 GNANTTTGGAAATNTTGGTNTTNTTGGATCTGNTTAGTTAGTCAN 750
251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
|||||
751 ACACANNAANNCACAGANNAATNCCNNTATGAAAGGANTANATCT 800
267 uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTrpIleP 284
|||||
801 TTCATCATTAANAATCAGCTNTCTCACAGAAGTCAGTCCGCTGNTAC 850
284 roMetLeuLysSerLeuProLeuTrpAlaIleValAlaHisPheSer 300
|||||
851 CNATNTNNAATCTGCCACTTTGGGCTATNGTNGTGCACANTTTTCT 900
301 TyrAsnTrpThrPheThrLeuLeuLeuLeuProThrTyrMetIly 317
|||||
901 TACAACCTGGACTTTTATACCTTTTACNTTATGCTACTTANATGAA 950
317 sGluValLeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValP 334
|||||
951 GGANNTCTAAGGTTCAATNTTCAAGAGAATGGGCTTTTATCTNCANTNC 1000
334 roTyrLeuGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAsp 350
|||||
1001 CTTATTTAGGNTNTGGTTTATGATGATCTCTGTCNGGCTCAAGCTGCAC 1050
351 AsnLeuArgAlaArgTrpAsnPheSerThrLeuTrpValArgValPhe 367
|||||
1051 AATTTAGGGCAANATGGAATTTTCAACTNTNGTTCGNAGANTTTT 1100
367 eSerLeuIleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyP 384
|||||
1101 TAGCCTTATAGGNATGATGGACCTGCNNTATTCCTGTCGTCGTCGTCG 1150
384 heIleGlyCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThr 400
|||||
1151 TNATNGGCTGCTGATATATNTTGGCNGTTCCTTCAACNATATCAACA 1200
401 ThrLeuGlyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspI 417
|||||
1201 ACNCTGGAGGCTTTGCTCTCTCGGATTTAGCATCAACCATCTGGANAT 1250
417 eAlaProSerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaT 434

|||||
1251 TGCCTCTCGTATGCTGGTATNCTCTGGGCATCACAAATACNTTTGCCA 1300
434 hrIleProGlyMetIleGlyProIleIleAlaArgSerLeuThrProGlu 450
|||||
1301 CTATTCNGGAATGTTGGGCCCTTCATTCGNANAAGCTTNACCCCTGAN 1350
451 AsnThrIleGlyGluTrpGlnThrValPheCysIleAlaAlaIleAs 467
|||||
1351 AACACTTTGGAGAAATGGCAACNGTNTCTNNATNGCTGCTGCTATNAA 1400
467 nValPheGlyAlaIlePhePheThrLeuPheAlaLysGlyGluValGlnA 484
|||||
1401 TGTNTTTGGTGCATTTTCTTNACACTATTGCCCAAAGGTGAAGTNCAAA 1450
484 snTrpAlaIleSerAspHisGlnGlyHisArg 494
|||||
1451 ACTGGCNCNTCANTGATCACCANGACACAGA 1482

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAH79234

seq_documentation_block:

ID AAH79234 standard; cDNA; 2670 BP.

XX AAH79234;
XX AC
XX 20-NOV-2001 (first entry)
XX Human sodium dependent phosphate co-transfer protein 35 cDNA.
DE Human; sodium dependent phosphate co-transfer protein 35;
XX KW hypophosphaturia; hypercalcaemia; hypophosphataemic rickets;
KW nephritis; gene therapy; ss.
XX OS Homo sapiens.
XX CN1298882-A.
XX PD 13-JUN-2001.
XX PF 06-DEC-1999; 99CN-0124217.
XX PR 06-DEC-1999; 99CN-0124217.
XX (SHEN-) SHENGYUAN GENE DEV CO LTD SHANGHAI.
XX PI Mao Y, Xie Y;
XX WPI; 2001-503367/56.
XX P-PSDB; AAG65238.
XX Human Na-dependent phosphate cotransporter 35 and its coding sequence -
PS Claim 6; Page 19-20(Disclosure); 28pp; Chinese.
XX The present invention provides the protein and coding sequences of human
CC sodium dependent phosphate co-transfer protein 35. The sequences can be
CC used in the treatment of hypophosphaturia, hypercalcaemia,
CC hypophosphataemic rickets and nephritis. The present sequence is the
CC coding sequence of the invention.
XX Sequence 2670 BP; 706 A; 550 C; 552 G; 862 T; 0 other;

alignment_scores:

Quality: 1578.00 Length: 321
Ratio: 5.042 Gaps: 0
Percent Similarity: 97.508 Percent Identity: 89.097

alignment_block:

US-09-776-865-4 x AAH79234 ..

Align seg 1/1 to: AAH79234 from: 1 to: 2670

175 GluGlyValThrTyrProAlaMetHisAlaMetTrpSerSerTrpAlaPr 191
|||||
3 GAGGGTGTACATTCCAGCCATCGATGCCATGGTCTCTTGGGCTCC 52

191 OProLeuGluArgSerLysLeuLeuSerIleSerTyrAlaGlyAlaGlnL 208
|||||
53 CCCTCTGTAAGAAAGCAAACTCTTAGCATTTTCATATGCAGAGCACAGC 102

208 euGlyThrValValSerLeuProLeuSerGlyValIleCysTyrTrpMet 224
|||||
103 TTGGGACAGTAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTATG 152

225 AsnTrpThrTyrValPheTyrPheGlyIleValGlyIleIleIleTrpPh 241
|||||
153 AATTGGACTTATGCTCTCTACTTTTGGTACTATTGGAATATTTGGTT 202

241 eIleLeuTrpIleCysLeuValSerAspThrProGluThrHisLysThrI 258
|||||
203 TCTTTTGTGGATCTGGTTAGTTAGTACACACACACACACACACACAGAA 252

258 leThrProTyrGluLysGluTyrIleLeuSerSerLeuLysAsnGlnLeu 274
|||||
253 TTCCCATTAAGAAAGGAATACATTTCTTTCATCATTAAGAAATCAGCTT 302

275 SerSerGlnLysSerValProTrpIleProMetLeuLysSerLeuProLe 291
|||||
303 TCTTCACAGAAGTCAGTGGCGTGGTACCCATTTTAAATCCCTGCCACT 352

291 uTrpAlaIleValAlaHisPheSerTyrAsnTrpThrPheTyrThrL 308
|||||
353 TTGGGCTATCGTAGTGCACACTTTTCTTACAACTGGACTTTTATATACTT 402

308 euLeuThrLeuLeuProThrTyrMetLysGluValLeuArgPheAsnIle 324
|||||
403 TATTGACATTTATGCTTACTTATGAAGGAGATCTTAAGGTTCAATGTT 452

325 GlnLysAsnGlyPheLeuSerAlaValProTyrLeuGlyCysTrpLeuCy 341
|||||
453 CAAGAGAATGGGTTTTATCTTCATTCATTCATTTAGGCTCTTGGTTATG 502

341 sMetIleLeuSerGlyGlnAlaAlaAspAsnLeuArgAlaArTrpAsnP 358
|||||
503 TATGATCCTGTCTGGTCAAGCTCTGCACATTTTAAAGGGCAAAATGGAAT 552

358 heSerThrLeuTrpValArgValPheSerLeuIleGlyMetIleGly 374
|||||
553 TTTCAACTTTATGTTGCGCAGAAATTTTAGCCCTTATAGGAATGATTGGA 602

375 ProAlaIlePheLeuValAlaAlaGlyPheIleGlyCysAspTrpSerLe 391
|||||
603 CTGTCAGTATTCCTGTAGTGTGGCTTCATTTGGCTGTGATTATCTTT 652

391 uAlaValAlaPheLeuThrIleSerThrThrLeuGlyGlyPheCysSerS 408
|||||
653 GGCCGTTGCTTTCTTAACATATATCAACACACACACACACACACACAC 702

408 erGlyPheSerIleAsnHisLeuAspIleAlaProSerTyrAlaGlyIle 424
|||||
703 CTGGATTTAGCATCAACATCTCGATTTAGTCTCTCGTATGCTGGTATC 752

425 LeuLeuGlyIleThrAsnThrPheAlaThrIleProGlyMetIleGlyPr 441
|||||
753 CTCTCGGCATCAACAATACATTTGGCCTATTTCCAGGAATGGTTGGCC 802

441 oIleIleAlaArgSerLeuThrProGluAsnThrIleGlyGluTrpGlnT 458
|||||
803 CGTCATTGTAAAGTCTGACCCCTGATAACACACTGTTGGAGAATGCAAA 852

458 hrValPheCysIleAlaAlaAlaIleAsnValPheGlyAlaIlePhePhe 474
|||||
853 CCGTGTCTTATATGCTGTGCTATTAAATGTTTGGTGCCATTTCTTT 902

475 ThrLeuPheAlaLysGlyGluValGlnAsnTrpAlaIleSerAspHisG1 491
|||||
903 ACACATTCCCAAGGTGAAGTACAAACTGGCTCTCAATCATCACCA 952

491 nGlyHisArgAsn 495
:|||||
953 TGGACACAGACAC 965

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAI59901

seq_documentation_block:
ID AAI59901 standard; cDNA; 929 BP.
XX
AC AAI59901;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 3890.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
(HYSE-) HYSEQ INC.
XX
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
WPI: 2001-442253/47.
DR P-PSDB; AAM40745.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 3890; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX

SQ Sequence 929 BP; 214 A; 224 C; 223 G; 268 T; 0 other;

alignment_scores:

Quality: 1217.00 Length: 270
Ratio: 4.810 Gaps: 0
Percent Similarity: 93.704 Percent Identity: 82.963

alignment_block:

US-09-776-865-4 x AAI59901

Align seg 1/1 to: AAI59901 from: 1 to: 929

1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGlyLeu 17
119 ATGAGGTCTCGGTTCGAGACCTGCCCGGACGATGGCAGGAGAGCAC 168
17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaPro 34
169 GGACCGCAGCGCTCTTCTACGGCGCCGCCACGGCGGAGCGCTCCAG 218
34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyPhe 50
219 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 268
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
269 TTCATTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 318
67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
319 GGTAGATTCAATCAACTTTAGAGATAATAGAACTTCCAAGCTGCCGTC 368
84 laGluHisSerAlaProLeuLysValLeuHisAsnGlnThrGlyLysLys 100
369 CAGAGCATTTCTGCTCCCAAAAGTTTCATCATATCAATCAAGCGGTAA 418
101 TyrArgTrpAspAlaGluThrGlnGlyTrpPheLeuGlySerPhePheTy 117
419 TACCAATGGGATGAGAACTCAAGGATGATGATGATGATGATGATGATGAT 468
117 rGlyTyrIleIlePheGlnProGlyTyrValAlaSerArgSerG 134
469 TGGCTACATCATCACAGATTCCTGGAGGATATGTTGCCAGCAAAATAG 518
134 lyGlyLysLeuLeuGlyPheGlyPheAlaThrAlaIlePheThr 150
519 GGGGAAATGCTGCTAGGATTTGGGATCCTTGGCACTGCTGCTCCTCACC 568
151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
569 CTGTTCACTCCCATGCTGAGATTTAGGAGTTGGACCACTCAATGTA 618
167 uArgAlaLeuGluGlyLeuGlyGlyValThrProAlaMethisA 184
619 CAGAGCACTAGAAGGACTAGGAGGGTGTACATTTCCAGCCATGCATG 668
184 laMetTrpSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
669 CCATGTGCTCTTCTTGGGCTCCCGCTTTGAAAGAGCAAACTCTTAGC 718
201 lIleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
719 ATTTCAATATCAGGAGCACAGCTTGGACAGTAATTTCTTCTCTCTTC 768
217 rGlyValIleCysTyrTrpMetAsnTrpThrTyrValPheTyrPhePheG 234
769 TGAATAAATTTGCTACTATATGAATGAGCTTATGCTTCTTCTTCTTTT 818
234 lylleValGlyIleIleTrpPheIleLeuTrpIleCysLeuValSerAsp 250
819 GTACTATGGAATATTTTGTCTTCTTTTGTGGATCTGGTTAGTAGTGAC 868

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAK93901

seq_documentation_block:

ID AAK93901 standard; cDNA; 853 BP.

XX
AC AAK93901;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA clone representative sequence, SEQ ID NO: 2361.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.

XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
DR WPI; 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Example 11; SEQ ID NO 2361; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence was used as the
CC representative sequence from a human clone which was used in
CC homology searches to identify the clone.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 853 BP; 198 A; 200 C; 213 G; 239 T; 3 other;

alignment_scores:

Quality: 1104.00 Length: 248
Ratio: 4.779 Gaps: 0
Percent Similarity: 93.145 Percent Identity: 82.258

alignment_block:

US-09-776-865-4 x AAK93901

Align seg 1/1 to: AAK93901 from: 1 to: 853

```
1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGlySe 17
||||:|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
83 ATGAGTCTCCGGTTCGAGACCTGGCCCGAACGATGGCGAGGAGAC 132

17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProV 34
:|||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||
133 GGACCGCAGCCTCTCTACCGGGCCGCCACGGCCGAAGCGCTCCAG 182

34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyPhe 50
||||:|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
183 TGTGCTGCTCTGCTCTTACCACTTAGCAATTTGGCCTTTTGGTTTC 232

51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
||||:|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
233 TTCATTGTGTATGCATTCAGTCTGATCTCAGTGTTCGTTAGTGGATAT 282

67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
|||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||
283 GGTAGATTCAAATACNACTTTAGAGATTAATAGAACTTCCAAAGCGGTGC 332

84 laGluHisSerAlaProLysValLeuHisAsnGlnThrGlyLysLys 100
|||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||
333 CAGAGCATTCCTCCCAATAAAGTTTCATCATATAATCAAACGGGTAAAGAAG 382

101 TyrArgTyrAspAlaGluThrGlnGlyTyrPheLeuGlySerPhePheTyr 117
||||:|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
383 TACCAATGGATGCAAACTCAAGATGGATTCGTTCTCTCTTTTAA 432

117 rGlyTyrIleThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134
|||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||
433 TGGCTACATCATCACAGATTCCTGGAGATATGTTGCCACCAAAATAG 482

134 lyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
||||:|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
483 GGGGGAATGCTGCTAGGATTTGGATCTTGGCACTGCTGCTCCAC 532

151 LeuPheThrProLeuAlaAlaPheGlyValGlyAlaLeuValAlaLe 167
|||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||
533 CTGTTCACTCCCATTCGTCAGATTTAGGATTTGGACCACTCATTTGACT 582

167 uArgAlaLeuGlyLeuGlyGlyValThrTyrProAlaMetHisA 184
|||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||
583 CAGAGCACTAGAGGACTAGAGAGGCTTTACATTTCCAGCCATGCATG 532

184 laMetTrpSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200
|||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||
633 CCATGTGGTCTTCTGGGCTCCCTCTTGAAGAAGCAAACTTCTTACG 682

201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
|||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||
683 ATTTCAATATGCAGGACACAGCTTGGCAGAGTAATTTCTTCTCTTTC 732

217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
|||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||  ||||||
733 TGAATAATTTGCTACTATATGCAATGGACTATGCTTCTTCTACTTTT 782

234 lyIleValGlyIleIlePhePheIleLeuTrpIleCysLeuVal 248
|||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||  ||||
783 GNACTATTGGAAAAATTTGGTCTCTTTTNGGATCTCTNGTAGTA 826
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK92364

seq_documentation_block:

ID AAK92364 standard; cDNA; 838 BP.

XX AC AAK92364;

XX DF 06-NOV-2001 (first entry)

XX DE Human cDNA 5'-end sequence, SEQ ID NO: 824.

XX KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX

```
OS Homo sapiens.
XX
PN EPI130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -
PS Claim 2; SEQ ID NO 824; 1380pp + sequence listing; English.
XX
XX The invention relates to primers for synthesising full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX molecules have been determined. Primers for synthesising the full length
XX cDNA are useful for clarifying the function of the protein encoded by
XX the cDNA. The full length clones were obtained by construction of full
XX length enriched cDNA libraries that were synthesised by the oligo-capping
XX method. The primers enable the production of the full length cDNA easily
XX without any special methods. The present sequence is the nucleotide
XX sequence of the 5'-end of a cDNA provided in the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 838 BP; 189 A; 196 C; 210 G; 237 T; 6 other;
```

alignment_scores:

Quality: 1022.50 Length: 244

Ratio: 4.585 Gaps: 1

Percent Similarity: 91.393 Percent Identity: 80.328

alignment_block:

US-09-776-865-4 x AAK92364

Align seg 1/1 to: AAK92364 from: 1 to: 838

1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGlySe 17

||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

83 ATGAGTCTCCGGTTCGAGACCTGGCCCGAACGATGGCGAGGAGAC 132

17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProV 34

:||||| |||||| |||||| |||||| |||||| |||||| |||||| |||||| |||||| ||||||

133 GGACCGCAGCCTCTCTACCGGGCCGCCACGGCCGAAGCGCTCCAG 182

34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyPhe 50

||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

183 TGTGCTGCTCTGCTCTTACAACTTAGCAATTTGGCCTTTTGGTTTC 232

51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67

||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

233 TTCATTGTGTATGCATTCAGTCTGATCTCAGTGTTCGTTAGTGGATAT 282

67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84

||||| |||||| |||||| |||||| |||||| |||||| |||||| |||||| |||||| ||||||

283 GGTAGATTCAAATACAACTTTAGAGATTAATAGAACTTCCAAAGCGGTGC 332

84 laGluHisSerAlaProLysValLeuHisAsnGlnThrGlyLysLys 100

||||| |||||| |||||| |||||| |||||| |||||| |||||| |||||| |||||| ||||||

333 CAGAGCATTCCTCCCAATAAAGTTTCATCATATAATCAAACGGGTAAAGAAG 382

```
101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheThy 117
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
383 TACCAATGGGATGCAGAACTCAAGATGATTCGGTTCCTTTTAA 432
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
117 rGlyTyrIleThrGlnIleProGlyGlyTyrValAlaSerArgSerg 134
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
433 TGGCTACATCATCACAGATTCCTGGAGGATATGTTGCCAGCAAAATAG 482
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
134 lYGLyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
483 GGGGAAATAGCTGCTAGGATTTGGGATCCTGGCACTGCTGCCACACC 532
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
533 CTGTTCACTCCCATGCTGCAGATTTAGGAGTTGGACCACCTATTGTACT 582
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
167 uArgAlaLeuGluGlyLeuGlyGluGlyValThrTyrProAlaMethHisA 184
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
583 CAGAGCACTAGAAGGACTANGAGAGGGTGTACATTTCCAGCCATGCATG 632
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
184 laMetTrpSerSerTrpAlaProProLeuGluuArgSerLysLeuLeuSer 200
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
633 CCATGTGCTCTCTTGGGCTCCCCCTCTTGAAGAAGCAAACTCTTAGC 682
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
683 ATTTCATATGCAGGACACANCTTGGGACAGTAATTTCTTNCCTCTTC 732
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
217 rGlyValIleCysTyrTyrMetAsnTrpThrTyrValPheTyrPhePheG 234
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
733 TGGAAATAAT.TGCTACTATATGAATGGGACTATGCTTCTACTTTTGG 781
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
234 lYleValGlyIleIleTrpPheIleLeuTrp 244
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
782 GGACTATGGAATATTGGGCTC...FTTTGG 810
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL03769
seq_documentation_block:
ID ABL03769 standard; cDNA; 1939 BP.
AC ABL03769;
XX
XX
XX 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 5789.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
PR
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
DR
XX
XX P-PSDB; ABB59666.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
```

```
PS Claim 1; SEQ ID NO 5789; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1939 BP; 410 A; 558 C; 552 G; 419 T; 0 other;
```

```
alignment_scores:
Quality: 991.50 Length: 517
Ratio: 2.857 Gaps: 7
Percent Similarity: 67.118 Percent Identity: 39.458
alignment_block:
US-09-776-865-4 x ABL03769 ..
Align seg 1/1 to: ABL03769 from: 1 to: 1939
11 SerAspGlyGluGlySerAspArg.....Th 20
||||| ||| :|||||
353 TCCGAGCAGCAGGAGCGACGAGGAGGCGTTCGTCGCGGCAACG 402
||||| ||| :|||||
20 rProLeuLeuGlnArgAlaProArgAlaGluProAlaProValCys...C 36
||||| :||| :|||
403 ACCGCTAATCCGCTCCAGCGCGCTCGGAGGAGAAACCATGGCTGGCC 452
||||| :||| :|||
36 ysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyPheVal 52
:||||:||||:||||:||||:||||:||||:||||:
453 CAAAGACGGCTCACATTTTCGGATTTCGGCTTCCTGGGATTTCGCGTG 502
||||| :||| :|||
53 LeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMetValAs 69
:||||:||||:||||:||||:||||:||||:||||:
503 GTCACGCGATCGGGTCAATCTGTCGGTGGCCATTTGGCCATGGTG... 550
||||| :||| :|||
69 pSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysAlaGluH 86
||||| :||| :|||
551 ....AACCAACGGCAATTCGCGACAGCACTCATCGGTGATTGATACG 596
||||| :||| :|||
86 isSerAlaProIleLysValLeuHisAsn..... 95
:||||:||||:||||:||||:||||:||||:||||:
597 ACACGTGTCCACTACCGCGCACCATCACATCACAATGGTAGCGATCCCAATCCG 646
||||| :||| :|||
96 GlnThrGlyLysLysTyrArgTrpAspAlaGluThrGlnGlyTrpIleLe 112
||||| :||| :|||
647 CAGAAGGAGGGGAGTGTGTGGGACGAGCGCACGAGGATTTGGTGT 696
||||| :||| :|||
112 uGlySerPhePheTyrGlyTyrIleIleThrGlnIleProGlyGlyTyrV 129
||||| :||| :|||
697 CGGCAGTTTCTTCTATGGCTATGCTACCCAAGTGCCCGGGGACGGA 746
||||| :||| :|||
129 alAlaSerArgSerGlyLysLeuLeuGlyPheGlyIlePheAla 145
:||||:||||:||||:||||:||||:||||:||||:
747 TGGCCGAGCTGTATGTTGGGAAGATCTACGGCTATGAGTGTGTGATC 796
||||| :||| :|||
146 ThrAlaIlePheThrLeuPheThrProLeuAlaAlaAspPheGlyValGl 162
||||| :||| :|||
797 ACGGCGGTCTTTACGCTTATACCTCATTCGTCGCTGCTGATGATCGC 846
||||| :||| :|||
162 yAlaLeuValAlaLeuArgAlaLeuGluGlyLeuGlyGluGlyValThrT 179
||||| :||| :|||
847 GCTGTTGGTCTTGGTCCGATCTGCGAGGAAATGGCGAGGGCGTCACCT 896
||||| :||| :|||
179 yProAlaMethHisAlaMetTrpSerTrpAlaProProLeuGluArg 195
||||| :||| :|||
```


OM of: US-09-776-865-4 to: EST:* out_format : pfs
Date: Jul 15, 2002 2:32 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cgn2.1/USPTO.spoal/US09776865/runat_15072002.082207.27382/app_query.fasta_1.1159
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LGOPEL=0.000 -LGOPEXT=0.000 -CGAPOP=4.500
-CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosom62 -TRANS=human40.cdi
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09776865.ecgn1.1.3468
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPY -WAIT -THREADS=1

Search information block:

Query: US-09-776-865-4
Query length: 495
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 2434.030000

score_list:

| Sequence | Strd Orig | ZScore | EScore | Len | Documentation |
|-------------------|-----------|---------|----------|------|-----------------------------------|
| gb_est1:AL550137 | + 1395.00 | 2364.64 | 1.8e-122 | 985 | ! AL550137 LTI_NFL006.PL2 |
| gb_est2:BF661062 | + 1038.00 | 1756.57 | 1.4e-88 | 801 | ! B1661062 603304362E1 NIH.CGAP.M |
| gb_est2:BF1907284 | + 953.00 | 1611.92 | 1.5e-80 | 754 | ! B1907284 603063858F1 NIH.MGC.11 |
| gb_est2:BF676817 | + 921.00 | 1559.22 | 1.3e-77 | 626 | ! BF676817 602084380F1 NIH.MGC.83 |
| gb_est2:BF67765 | + 915.00 | 1547.59 | 5.9e-77 | 711 | ! BF67765 603346858F1 NCI.CGAP.M |
| gb_hic:BC023331 | - 900.00 | 1513.67 | 4.6e-75 | 1531 | ! BC023331 Mus musculus, clone I |
| gb_est2:BF659521 | + 879.50 | 1486.53 | 1.5e-73 | 736 | ! B1860521 603386787F1 NIH.MGC.87 |
| gb_est2:BF539146 | + 868.00 | 1464.76 | 2.4e-72 | 895 | ! B539146 602053009F1 NCI.CGAP.9 |
| gb_est1:BF537525 | + 839.50 | 1419.57 | 8.0e-70 | 645 | ! B537525 B537325 RIKEN full-le |
| gb_est2:BF124137 | + 837.00 | 1411.65 | 2.2e-69 | 904 | ! BF124137 601760838F1 NCI.CGAP.M |
| gb_est1:BF613552 | + 821.00 | 1387.34 | 5.0e-68 | 682 | ! B613552 B613352 RIKEN full-le |
| gb_est2:BF6541099 | + 808.00 | 1362.21 | 1.3e-66 | 893 | ! B6541099 602570265F1 NIH.MGC.77 |
| gb_est1:AA833297 | + 803.00 | 1359.54 | 1.8e-66 | 518 | ! AA833297 ud05d09.r1 Soares.NMPU |
| gb_est2:BF1817031 | + 796.00 | 1345.41 | 1.1e-65 | 633 | ! B1817031 UMN07B04 Canine Brain |
| gb_est2:BF867611 | + 782.00 | 1320.77 | 2.5e-64 | 676 | ! B867611 601443127F1 NIH.MGC.65 |
| gb_est2:BF869819 | + 736.00 | 1241.75 | 6.4e-60 | 700 | ! B869819 60446651F1 NIH.MGC.65 |
| gb_est2:BF181890 | + 716.50 | 1209.09 | 4.2e-58 | 658 | ! B181890 603378047F1 NCI.CGAP.M |
| gb_est2:BF866603 | + 709.00 | 1190.72 | 4.5e-57 | 1100 | ! B866603 602785577F1 NCI.CGAP.M |
| gb_est1:BF610013 | + 698.00 | 1177.93 | 2.3e-56 | 630 | ! B610013 B610013 RIKEN full-le |
| gb_est1:BF664731 | + 697.50 | 1177.09 | 2.6e-56 | 629 | ! B664731 B664731 RIKEN full-le |
| gb_est2:BF179717 | + 694.50 | 1172.71 | 4.5e-56 | 587 | ! B179717 dal11a10.y1 NICHD.XC |
| gb_est2:BF060500 | + 673.00 | 1133.73 | 6.7e-54 | 721 | ! B060500 BU060500 NIBB Mochii r |
| gb_est2:BF563945 | + 662.00 | 1120.53 | 3.6e-53 | 429 | ! B563945 BU060500 NIBB Mochii r |
| gb_est1:AA258513 | + 638.00 | 1079.43 | 7.1e-51 | 432 | ! AA258513 zr59d01.r1 Soares.NHMH |
| gb_est2:BF031727 | + 624.00 | 1051.39 | 2.6e-49 | 632 | ! B031727 B031727 NIBB Mochii r |
| gb_est1:AL597124 | + 576.00 | 971.88 | 6.9e-45 | 499 | ! AL597124 DKEP2313H91F1 r1.313 |
| gb_est2:BF375421 | + 538.00 | 908.16 | 2.4e-41 | 445 | ! B375421 601230485F1 NCI.CGAP.M |
| gb_est2:BF445533 | + 529.00 | 886.73 | 3.0e-40 | 647 | ! B445533 dae11h09.y3 NICHD.XC |
| gb_est2:BF499662 | + 529.00 | 886.29 | 4.0e-40 | 811 | ! B499662 AT14282.5prime AT Dros |
| gb_gss:CN50069V | + 506.00 | 843.93 | 9.2e-38 | 1075 | ! AL063986 Drosophila melanogast |
| gb_est2:BF123545 | + 504.00 | 842.85 | 1.1e-37 | 866 | ! B123545 601759862F1 NIH.MGC.94 |
| gb_est2:BF123545 | + 500.50 | 838.10 | 2.0e-37 | 772 | ! B123545 601759862F1 NCI.CGAP.M |
| gb_est2:BF82556 | + 497.00 | 832.29 | 4.1e-37 | 760 | ! B82556 601433358F1 NIH.MGC.72 |
| gb_est2:BF3685 | + 494.50 | 835.87 | 2.6e-37 | 367 | ! B3685 yr55601.r1 Soares fetal |
| gb_est2:BF134022 | + 481.50 | 805.89 | 1.2e-35 | 753 | ! B134022 60335557F1 NIH.MGC.94 |
| gb_gss:CN50352N | + 475.00 | 792.35 | 6.9e-35 | 943 | ! AL229388 Tetraodon nigroviridis |
| gb_est2:BF1753164 | - 465.50 | 779.95 | 3.4e-34 | 661 | ! B1753164 603026028F1 NIH.MGC.11 |
| gb_est2:BF495471 | + 465.50 | 779.65 | 3.5e-34 | 679 | ! B495471 AT04410.5prime AT Dros |
| gb_est2:BF12354 | + 453.00 | 767.07 | 1.8e-33 | 301 | ! B12354 yr53H06.r1 Soares meland |
| gb_est2:BF502767 | + 435.50 | 727.93 | 2.7e-31 | 707 | ! BF502767 AT18580.5prime AT Dros |

gb_est2:BE977644 + 433.50 725.67 3.6e-31 635 ! BE977644 bs67c07.y1 Drosophi
gb_est2:BF505525 + 431.50 719.95 7.4e-31 786 ! BF505525 AT07766.5prime AT D
gb_est1:AW531337 - 431.00 728.77 2.4e-31 321 ! AW531337 UI-R-C4-akw-c-03-0-
gb_est2:BG296175 + 430.00 717.43 1.0e-30 783 ! BG296175 602394006F1 NIH.MGC
gb_est2:BG123873 + 427.50 713.46 1.7e-30 761 ! BG123873 ES1469519 tomato sh
seq_name: gb_est1:AL550137
seq_documentation_block:
LOCUS AL550137 985 bp mRNA linear EST 16-FEB-2001
DEFINITION AL550137 LTI_NFL006.PL2 Homo sapiens cDNA clone CSOD1040YK17.5
prime, mRNA sequence.
ACCESSION AL550137
VERSION AL550137.1 GI:128868913
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 985)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
Source
1..985
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="LTI_NFL006.PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fulllength.invitrogen.com
http://fulllength.invitrogen.com"
BASE COUNT 236 a 224 c 224 g 300 t
ORIGIN
alignment_scores:
Quality: 1395.00 Length: 309
Ratio: 4.777 Gaps: 0
Percent Similarity: 94.498 Percent Identity: 83.819
alignment_block:
US-09-776-865-4 x AL550137 ..
Align seg 1/1 to: AL550137 from: 1 to: 985
1 MetLysSerProLeuSerAspLeuAlaProSerAspGlyGluGluGlyse 17
60 ATGAGGTCCTCGGTCGAGACCTGGCCGGACGACGAGGAGGAGCAC 109
17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProV 34
110 GGACCGGACCGCCCTCTCTACCGGGCGGCCCGGCGGACCGGCTCCAG 159
34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPheGlyphe 50
160 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTTC 209
51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67
210 TTCAATGCTGATGCATACGTGGAATCTGAGTGTGCTGCTGCTGCTGCTG 259

```
67 tValaspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84
|||||
260 GGTAGATTCAATCACTTTAGAGATAATAGAACATCCCAAGCGGTCTC 309
|||||
84 laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100
|||||
310 CAGAGCATCTCGTCCCAATAAAGTTCATCATATAATCAACGCGGTAAGA 359
|||||
101 TyrArgTyrAspAlaGluThrGlnGlyTyrIleLeuGlySerPhePhe 117
|||||
360 TACCAATGGGATGCAGAAACCTCAAGATGGATCTCGGTCTCTTTTATA 409
|||||
117 rGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134
|||||
410 TGGGTACATCATCACAGATCTCTGGAGGATATGTTGCCAGCAATA.G 458
|||||
134 lyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150
|||||
459 GGGGAAATGCTGCTAGGATTTGGATCCTTGGCACTGCTGCTCACC 508
|||||
151 LeupheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167
|||||
509 CTGTTCACTCCATGTGTCAGATTTAGGAGTTGGACCACTCATTTGACT 558
|||||
167 uArgAlaLeuGluGlyLeuGlyGlyValThrTyrProAlaMethHisA 184
|||||
559 CAGAGCACTAGAGAGCTAGAGAGGGTGTACATTTCCAGCCATGTCATG 608
|||||
184 laMetTyrSerSerTyrAlaProProLeuGluArgSerLysLeuLeuSer 200
|||||
609 CCATGTGCTCTCTTGGGCTCCCTCTTGAAGAGCAAACTCTTAGC 658
|||||
201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValValSerLeuProLeuSe 217
|||||
659 ATTTTCATATGAGAGGACAGCTTGGGACAGTAATTTCTTCTCTCTTTC 708
|||||
217 rGlyValIleCysTyrTyrMetAsnTyrThrTyrValPheTyrPhePhe 234
|||||
709 TGGATATATTTGCTACTATATGATGAGTATGCTCTTCTACTCTTTTGG 758
|||||
234 lyIleValGlyIleIleThrPheIleLeuThrPheIleCysLeuValSer 250
|||||
759 GTACTATTTGAATATTTTGGTTCTTTTGTGGATCTGTTAGTTAGTCAC 808
|||||
251 ThrProGluThrHisLysThrIleThrProTyrGluLysGluTyrIleLe 267
|||||
809 ACACCACAAAACACAGAGAAATTTCCCATTTATGAAAAGGAATACATTCT 858
|||||
267 uSerSerLeuLysAsnGlnLeuSerSerGlnLysSerValProTyrIleP 284
|||||
859 TTCATCATTAAGAATCAGCTTTCTTCCAGAAATCAGTGGCGGGGTAC 908
|||||
284 roMetLeuLysSerLeuProLeuTyrAlaIleValAlaHisPheSer 300
|||||
909 CCATTTTAAATCCTGCCACTTTGGGCTATCGTAGTTGCACACTTTTCT 958
|||||
301 TyrAsnTyrPheTyrThrLeuLeu 309
|||||
959 TACAACCTGGCTTTTATACTTTATTG 985
|||||
```

seq_name: gb_est2:BI661062

seq_documentation_block:
LOCUS BI661062 801 bp mRNA linear EST 12-SEP-2001
DEFINITION 603304362F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5350046 5',
mRNA sequence.
ACCESSION BI661062
VERSION BI661062.1 GI:15575298
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia: Eutheria: Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 801)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
Ph.D.

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11889 row: 1 column: 15
High quality sequence stop: 799.

FEATURES
source

Location/Qualifiers
1..801
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5350046"
/clone_lib="NIH_CGAP_Mam4"
/tissue.type="tumor, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1; NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert 2.5 kb. Library constructed by Life
Technologies, catalog # 12018-016. Investigators providing
samples: Lothar Hennighausen/Priscilla Furth, NIH
Reference for transgenic model: Li et al., Cell Growth and
Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP
Library."

BASE COUNT 173 a 200 c 220 g 208 t
ORIGIN

alignment_scores:

Quality: 1038.00 Length: 247
Ratio: 4.553 Gaps: 4
Percent Similarity: 92.308 Percent Identity: 81.377

alignment_block:

US-09-776-865-4 x BI661062

Align seg 1/1 to: BI661062 from: 1 to: 801

```
9 AlaProSerAspGlyGluGlySerAspArgThrProLeuLeuGlnAr 25
|||
52 GCGGAAACGACGATGAGAGAGCTCGACAGCACCCGCTCTGCCGGG 101
|||
25 gAlaProArgAlaGluProAlaProValCysSerAlaArgTyrAsnL 42
|||
102 CGCCCGGCAGACCGGCGCTCCAGTGTGCTGCTGCTGCGGTACAACT 151
|||
42 euAlaPheLeuSerPhePheGlyPheValLeuTyrSerLeuArgVal 58
|||||
152 TAGCGATTTTGGCGTCTCTGCGTTCTTCGTTCTCTATGCTTACGGGTG 201
|||||
59 AsnLeuSerValAlaLeuValAspMetValAspSerAsnThrThrAlaLy 75
|||||
202 AACCTGAGTGTGCTTAGTGGACATGCTAGATTCAAAATCAACTCTGAC 251
|||||
75 sAspAsnArgThrSerTyrGluCysAlaGluHisSerAlaProIleLys 92
|||||
252 TGATAATAGAACGCTAAGAGTGTGCGGAACATTTCTGCCCCATAAAG 301
|||||
92 alLeuHisAsnGlnThrGlyLysLysTyrArgTyrAspAlaGluThrGln 108
|||
302 TTCACCAATACACAGGTAAAGTACAGTGGGATGCGAGAACTCAA 351
|||||
109 GlyTyrIleLeuGlySerPheTyrGlyTyrIleIleThrGlnIlePr 125
|||||
```

```
352 GGTGGATTCTCGCTCTCTTTTACGGCTACATCTCACCAGATTCC 401
125 oglyGlyTyrValAlaSerArgSerGlyGlyLysLeuLeuGlyPheG 142
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
402 CGGTGGGTACATTGCCAGAGGTGCGAGGGAAGCTGCTGGCCCTGG 451
142 lyilePheAlaThrAlaIlePheThrLeuPheThrProLeuAlaAlaAsp 158
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
452 GCATCTTAGGCACCTCGCTTCCACCTGTTCACACGCTGGCGGCAGAC 501
159 PheGlyValGlyAlaLeuValAlaLeuArgAlaLeuGlyLeuGlyG1 175
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
502 TTAGCGGTGTGACTTCGTTGGCTTAGAGCGCTGAAGAGACTGGGAGA 551
175 uGlyValThrTyrProAlaMethHisAlaMetTrpSerSerTrpAlaProp 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
552 GGTGTTACGCTTCCAGCTATGCACGCCATGTGCTCTTCCTGGGCTCCC 601
192 roLeuGluArgSerLysLeuLeuSerLysSerTyrAlaGlyAlaGlnLeu 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
602 CTCTGGAAAGAACAGCTTCTTACCATTTCCTATGCGGAGCACAGCTT 651
209 GlyThrValValSerLeuProLeuSerGlyValIleCysTyrTyrMetAs 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
652 GGCAGACTGATCTCACTCTCTTTCGGAAATATATGCTACTATATGAA 701
225 nTrpThrTyrValPheTyr.PhePheGlyIleValGlyIleIleTrp.Ph 241
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
702 CTGGACTTACGCTCTCTATCTTT...GGTATAGTTGGAATTGCTGTTA 748
241 eileLeuTrpIleCysLeuValSerAspThrProGlu 253
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
749 T...TTATGGATGGGATAGTACGTACATACACCAGAA 782
```

seq_name: gb_est2:BI907284

seq_documentation_block: 754 bp mRNA linear EST 16-OCT-2001
LOCUS BI907284 603063858F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5212766 5',
DEFINITION mRNA sequence.

ACCESSION BI907284
VERSION BI907284.1 GI:16170094

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 754)

NIH-MGC <http://mgc.ncl.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@remail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLM11534 row: d column: 15

High quality sequence stop: 750.

Location/Qualifiers

1..754

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="5212766"

/clone_lib="NIH_MGC_118"

/tissue_type="Leukocyte"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV

(destroyed); RNA source leukocytes from anonymous pool of

non-activated adult donors. Library is oligo-dT primed

and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH_MGC Library."

BASE COUNT 171 a 182 c 194 g 207 t
ORIGIN

alignment_scores:

Quality: 953.00 Length: 227

Ratio: 4.453 Gaps: 1

Percent Similarity: 94.273 Percent Identity: 83.260

alignment_block:

US-09-776-865-4 x BI907284 ..

Align seg 1/1 to: BI907284 from: 1 to: 754

1 MetLysSerProValSerAspLeuAlaProSerAspGlyGluGluGlyse 17

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

78 ATGAGGTCTCCGGTTCGAGACCTGGCCCGGAACGATGCGAGGAGACAC 127

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

17 rAspArgThrProLeuLeuGlnArgAlaProArgAlaGluProAlaProV 34

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

128 GGACCG.ACGCCTCTTACCGGGCGCCACGGCGCGAAGCCGCTCCAG 176

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

34 alCysCysSerAlaArgTyrAsnLeuAlaPheLeuSerPhePheGlyPhe 50

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

177 TGTGCTGCTCTGCTCGTTACAACCTTAGCAATTTGGCCTTTTGGGTTTC 226

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

51 PheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuValAspMe 67

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

227 TTCATTGCTGATGCATTACGTTGAATCTGAGTTCGCTTAGTGATAT 276

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

67 tValAspSerAsnThrThrAlaLysAspAsnArgThrSerTyrGluCysA 84

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

277 GGTAGATTCAATACAACTTTAGAAGATAATAGAACTTCCAAGCGGTGTC 326

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

84 laGluHisSerAlaProIleLysValLeuHisAsnGlnThrGlyLysLys 100

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

327 CAGAGCATCTGCTCCCAATAAAGTTTCATCAATAACAAACGGGTAAAG 376

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

101 TyrArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerPhePheTy 117

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

377 TACCAATGGGATGCAGAACTCAAGGATGGATTCCTCGTTCCTTTTATA 426

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

117 rGlyTyrIleIleThrGlnIleProGlyGlyTyrValAlaSerArgSerG 134

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

427 TGGCTACATCATCACAGATTCTCGGAGGATATGTTGCCAGCAAAATAG 476

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

134 lyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIlePheThr 150

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

477 GGGGAAAATGCTGTAGGATGGGATCCTTGGCACTGCTGCTCCTCACC 526

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

151 LeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuValAlaLe 167

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

527 CTGTTCACTCCCAATGCTCGAGATTTAGGAGTTGGACCACTCATTTGTA 576

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

167 uArgAlaLeuGlyLeuGlyGluGlyValThrTyrProAlaMethHisA 184

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

577 CAGAGCACTAGAGGACTAGGAGGGGTGTACATTTCCAGCCATGCATG 626

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

184 laMetTrpSerTrpAlaProProLeuGluArgSerLysLeuLeuSer 200

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

627 CCATGTGGTCTCTTGGGCTCCGCCCTCTTGAAGAAGCAAACTTCTTAG 676

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

201 IleSerTyrAlaGlyAlaGlnLeuGlyThrValVal.SerLeuProLys 217

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

677 ATTTTCATATCGAGGAGCACAGCTTGGACAGTAGTAATTTTCTCTCTCT 726

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

217 erGlyValIleCysTyrTyrMetAsnTrp 226

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||


```

1230 GCAACTGGCGTGGCCATCGTGTCCATGGTCAACAACAGCACAACCCAC 1181
75  LysAspAsnArgThrSerTyrGluCysAlaGluHisSerAlaProIleIle 91
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1180 CGTGGGGCCCGAGTGGTGGTGAAGAACCCAG..... 1148
91  sValLeuHisAsnGlnThrGlyLysTyrArgTyrAspAlaGluThrG 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1147 .....TCAACTGGGATCCAGAGACTG 1126
108  InGlyTrpIleLeuGlySerPhePheTyrGlyTyrIleIleThrGlnIle 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1125 TCGGCTCATACATCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1076
125  ProGlyGlyTyrValAlaSerArgSerGlyGlyLysLeuLeuGlyPhe 141
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1075 CCTGAGGATTTATCGCAAAATTCGACGCAACAGGGCTTTGGCTT 1026
141  eGlyIlePheAlaThrAlaIlePheThrLeuPheThrProLeuAlaAla 158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1025 TGCCATTGTGCTACCTCCACCTAAACATGTGTGATCCCTTCAGCAGGCC 976
158  sPheGlyValGlyAlaLeuValAlaLeuArgAlaLeuGluGlyLeuGly 174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
975  CGGTTCACTATGGCTGTGTCTATCTTCGTGAGGATCCCTTCAGGATTTGGT 926
175  GluGlyValThrTyrProAlaMetHisAlaMetTrpSerSerTrpAlaPr 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
925  GAGGGGTACATACCCCTGCTGTGCATGGCATCTGGACAAATGGGCCCC 876
191  oProLeuGluArgSerLysLeuLeuSerIleSerTyrAlaGlyAlaGln 208
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
875  TCCCTTAGACGGAGTCGGTGGCAACACAGCACCTTTTGGGTTCTCTATG 826
208  euGlyThrValValSerLeuProLeuSerGlyValIleCysTyrTyrMet 224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
825  CTGGGGCGGTGGTGGCATGCGCTTGGCTGGGTCCTTGTGAGTATTCA 776
225  AsnTrpThrTyrValPheTyrPhePheGlyIleValGlyIleIleTrpPh 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
775  GGATGGAGTTCTGCTCTCTATGCTATGGCAGCTTCGGGATCTTTTGGTA 726
241  eIleLeuTrpIleCysLeuValSerAspThrProGluThrHisLysThrI 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
725  CCGTGTCTGGTGTCTGTCTCTATGATGATCACCAGCAGCTGCACCCAGCA 676
258  leThrProTyrGluLysGluTyrIleLeuSerSerLeuLysAsnGlnLeu 274
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
675  TCTCTGAGGAGGCGCAATACATTGAGGATGCCATCGGGGAGAGCGCC 626
275  SerSerGlnLysSerVal.....ProTrpIleProMetLe 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
625  AAGCTCATGACCTGTTCACGAAGTTTAACACACCTTGGAGCGCTCTT 576
286  uLysSerLeuProLeuTrpAlaIleValValAlaHisPheSerTyrAsn 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
575  TACGTCCATGCCCTCTATGTCATCATCGTTGCGAACTTTTGGCGCAGCT 526
303  rpThrPheTyrThrLeuLeuThrLeuLeuProThrTyrMetLysGluVal 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
525  GGACCTTCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 476
320  LeuArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValProTyrLe 336
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
475  TTCGGCTTGAGATCAGCAAGTGGGCTGTGTGCGGCTGCTCCTCAGCT 426
336  uGlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAspAsnLeu 353
   | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
425  TGTCTATGACCATCATCGTACCATTGGAGGCCAGATCGCTGACTTTTTCG 376
353  rgAlaArgTrpAsnPheSerThrLeuTrpValArgValPheSerLeu 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
375  CGAGTGGTGCATATGTCCTACCTACCAACGTCGCAAAAGCTCATGACGTC 326

```

```

370  IleGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyPheIleG 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
325  GGGGTTTCGGATGGAAGCCACGCTGCTGCTGGTGGTGGATAC...TC 279
386  yCysAspTyrSerLeuAlaValAlaPheLeuThrIleSerThrThrLeuG 403
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
278  GCATCCCAAGGGCGTGGCATCTCTCTCTGCTGCTGCTGGGCTTCA 229
403  lyGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIleAlaPro 419
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
228  GTGGCTTGGCCATCTCTGGTTTAACTGTAACCACTTGACATCGCCCT 179
420  SerTyrAlaGlyIleLeuLeuGlyIleThrAsnThrPheAlaThrIlePr 436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
178  CGCATGCCAGCATCTTGATGGCATTTCCCAATGGCGTGGGCACACTGTC 129
436  oGlyMetIleGlyProIleIleAlaArgSerLeuThrProGluAsnThrI 453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
128  TGGGATGGTGGCCCCATCATCGTGGTGCAATGACCAAGCACAAGACGC 79
453  leGlyGluTrpGlnThrValPheCysIleAlaAlaIleAsnValPhe 469
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
78  GGGAGGAGTGGCAGTACGTGTCTCTCATAGCTCCCTGGTGCACTACGC 29
470  GlyAlaIlePheThrLeuPheAla 478
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
28  GGTGTCATCTCTATGGGGTCTTTGCT 2

```

seq_name: gb_est2:BI860521

seq_documentation_block:

LOCUS BI860521 736 bp mRNA linear EST 10-OCT-2001
DEFINITION 60386787F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5395772 5',
mRNA sequence.

ACCESSION BI860521

VERSION BI860521.1 GI:16001268

KEYWORDS EST.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 736)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs@email.nih.gov

Tissue Procurement: DCTD/DTP

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM12008 row: m column: 21

High quality sequence stop: 713.

Location/Qualifiers

1..736

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5395772"

/clone_lib="NIH_MGC_87"

/tissue_type="mammary adenocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: breast; Vector: pCMV-SPORT6; Site:1: NotI;

Site:2: SalI; Cloned unidirectionally; oligo-dT primed.

Average insert size 1.383 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC library."

216 a 142 c 147 g 231 t

BASE COUNT

ORIGIN

```
alignment_scores:
  Quality: 879.50      Length: 224
  Ratio: 4.557         Gaps: 2
  Percent Similarity: 86.161      Percent Identity: 75.000

alignment_block:
US-09-776-865-4 x BI860521 ..
Align seg 1/1 to: BI860521 from: 1 to: 736

79 ThrSerTyrGluCysAlaGluHisSerAlaProIleLysValLeuHisAs 95
|||||
1 ACTTCCAAGCGGTGTCAGAGCACTCTGCTCCCATAAAGTTTCATCAATA 50
95 nGlnThrGlyLysLysTyrArgTgPAspAlaGluThrGlnGlyTrpIleL 112
|||||
51 TCRAACGGGTAAAGATACCAATGGGATGCAGAACTCAAGGATGGATTC 100
112 euGlySerPhePheTyrGlyTyrIleIleThrGlnIleProGlyGlyTyr 128
|||||
101 TCGGTTCCTTTTATATGGGTACATCATCACAGATTCCTGGAGGATAT 150
129 ValAlaSerArgSerGlyGlyLysLeuLeuGlyPheGlyIlePheAl 145
|||||
151 GTTCCAGCAAAATAGGGGGAAATGCTGCTAGGATTTGGGATCCTTGG 200
145 aThrAlaIlePheThrLeuPheThrProLeuAlaAlaAspPheGlyValG 162
|||||
201 CACTGCTGTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 250
162 lAlaLeuValAlaLeuAlaLeuAlaLeuGlyGlyGlyGlyValThr 178
|||||
251 GACCACCTATTGTACTCAGAGCACTAGAGGACTAGGAGGAGGTTTACA 300
179 TyrProAlaMetHisAlaMetTrpSerSerTrpAlaProProLeuGluAr 195
:::
301 TTTCCAGCCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG 350
195 gSerLysLeuLeuSerIleSerTyrAlaGlyAlaGlnLeuGlyThrValV 212
|||||
351 AGCAAACTCTTACGATTTTCATGATGAGGAGGAGGAGGAGGAGGAGTAA 400
212 alSerLeuProLeuSerGlyValIleCystTyrTyrMetAsnTrpThrTyr 228
:::
401 TTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 450
229 ValPheTyrPhePheGlyIleValGlyIleIleTrpPheIleLeuTrpI 245
|||||
451 GTCTTCTTACTT.TTTGGTACTATTGGAATATTTTGGTCTCTTTTGGAT 499
245 eCysLeuValSerAspThrProGluThrHisLysThrIleThrProTyrG 262
|
500 CTGGTTAGTTAGTACACACCAACAAACACAGAGAAATTTCCCATTTATG 549
262 lLysGluTyrIleLeuSerSerLeuLysAsnGlnLeuSerSerGlnLys 278
|||||
550 AAAAGGAATACATCTTTCATCATTAGAAATCAG..... 584
279 SerValProTrpIleProMetLeuLysSerLeuProLeuTrpAlaIleVa 295
:::
585 .....GTATGGAGTTG... 596

295 lValAlaHisPheSerTyrAsn 302
:::
597 ....ACACATTTTAAAAATAAT 614

seq_name: gb_est2:BF539146

seq_documentation_block:
LOCUS BF539146
DEFINITION 602053009F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4192035 5',
mRNA sequence.
```

```
|||||
356 GGGTGGATTCGGCTCTTTTGTACGGCTACATCGTCCAGATTC 405
|||||
125 roGlyGlyTyrValAlaSerArgSerGlyGlyLysLeuLeuGlyPhe 141
|||||
406 CGGTGGGTACATGCCAGAGGTCGGAGGAAGCTCTCTGGGCGTG 455
|||||
142 GlyIlePheAlaThrAlaIlePheThrLeuPheThrProLeuAlaAlaAs 158
|||||
456 GGCATCTTAGCCACCTCGCTTCACCCCTGTTACACCGCTGGCGCAGA 505
|||||
158 pPheGlyValGlyAlaLeuValAlaLeuArgAlaLeuGluGlyLeuGlyG 175
|||||
506 CTTAGGCGTGTGACTCTCTCGTGTG.CTTAGAGCGCTGGAAGACTGGGAG 554
|||||
175 LuGlyValThrTyrProAlaMetHisAlaMetTrpSerSerTrpAlaPr 191
|||||
555 AGGTGTACCTTCCAGCTATGCAGCGCATGTGCTCTCTGGGCTCC 604
|||||
191 oProLeuGluArgSerLysLeuSerIleSerTyrAlaGlyAlaGlnL 208
|||||
605 CCTCTGGAAGAAGCAAGCTTCTTACCATTCCTATTCGGGAGCACAGC 654
|||||
208 euGlyThrValValSerLeuProLeuSerGlyValIleCysTyrTyrMet 224
|||||
655 TTGGACAGTATCTACTCTCTCTGCGGAATAATATGCTACTATATG 704
|||||
225 AsnTrp.ThrTyrValPhe.TyrPhePheGlyIleValGlyIleIleTrp 240
|||||
705 AACTGGGACTTAAGTCTCTAAC...TTGGACTAGCGGAGTATGGCGGT 751
|||||
241 PheIleLeuTrpIleCysLeuValSerAspThrProGluThrHisLysTh 257
|||||
752 ...ACACAATGGGAGGGGAGCGCGTGATACCCCGAAATTCGAGAAA... 796
|||||
257 rIleThrProTyrGluLysGluTyrIleLeuSerSerLeuLysAsnGlnL 274
|||||
797 ...TTCCCTTTTGGAAAGACAC.....ACTGGTGCCTCTAAAAAACAGTT 836
|||||
274 euSerSerGlnLysSerValProTrpIlePro 284
|||||
837 TGGAAAGAGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 868
```

seq_name: gb_estl:BB537525

seq_documentation_block:

LOCUS BB537525 645 bp mRNA linear EST 26-OCT-2001
DEFINITION BB537525 RIKEN full-length enriched, 0 day neonate eyeball Mus
musculus cDNA clone E130006F23 3' similar to AJ387747 Homo sapiens
mRNA for sialin, mRNA sequence.

ACCESSION BB537525 GI:16446752

VERSION BB537525.2

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 645)

Akakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
, M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

On Jul 31, 2000 this sequence version replaced gi:9593025.

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,M., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

FEATURES

source

1..645
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="E130006F23"
/clone_lib="RIKEN full-length enriched, 0 day neonate
eyeball"
/tissue_type="eyeball"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: Sall; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCGCAACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCTCGAGTTAATTAATATCCCGCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 153 a 154 c 160 g 177 t
ORIGIN

alignment_scores:

Quality: 839.50 Length: 191
Ratio: 4.716 Gaps: 2

Percent Similarity: 93.194 Percent Identity: 83.770

alignment_block:

US-09-776-865-4 x BB537525 ..

Align seg 1/1 to: BB537525 from: 1 to: 645

49 GlyPhePheValLeuTyrSerLeuArgValAsnLeuSerValAlaLeuVa 65


```
|||||
20 GGGTCTCTGCTCTATGCTTANCGGTGAACCTGAGTGTGCGTAGT 69
65 lAspMetValAspSerAsnThrTrpAlaLysAspAsnArgThrSerTyrg 82
70 GGACATGGTAGATCAATACAACTCTGACTGATATAGAACGCTAAGG 119
82 luCysAlaGluHisSerAlaProIleLysValLeuHisAsnGlnThrGly 98
120 AGTGTGGGAACATCTGCCCCCATAAAAGTTCACCAACAATCACACAGT 169
99 LysLysTyrgArgTrpAspAlaGluThrGlnGlyTrpIleLeuGlySerph 115
170 AAAAAGTACAAAGTGGGATGAGAACTCAAGGGTGGATTCTCGGCTCTT 219
115 ePheTyrgLysTyrlleIleThrGlnIleProGlyGlyTyrgValAlaSerA 132
220 TTTTACGGCTACATCTACCCAGATCCCGGTGGGTGATATGCCAGCA 269
132 rgSerGlyGlyLysLeuLeuGlyPheGlyIlePheAlaThrAlaIle 148
270 GGGTCGGAGGAAAGCTGCTGCTGGCCCTGGGCATCTTAGCACCTCGTC 319
149 PheThrLeuPheThrProLeuAlaAlaAspPheGlyValGlyAlaLeuVa 165
320 TTCACCTGTCTACACCGCTGGCGCAGACTTAGGCGGTGAGTCTCGT 369
165 lAlaLeuArgAlaLeuGluGlyLeuGlyGlu.....Gly. 176
370 TGTGCTTAGAGCGCTGGAAGAGCTGGGAGAGAAATATCCACCCAGGT 419
177 ValThrTyrgProAlaMetHisAlaMetTrpSerSerTrpAlaProLe 193
420 GTTACGTTTCCAGTATGACGCCATGTGTCTTCCCTGGGCTCCCTCT 469
193 uGluArgSerLysLeuSerLysSerTyrgAlaGlyAlaGlnLeuGlyT 210
470 GGAAGAAGCAAGCTCTTACCAATTTCTATGCGGGAGCACAGCTGGGA 519
210 hrValValSerLeuProLeuSerGlyValIleCysTyrgTyrgMetAsnTrp 226
520 CAGTGATCTCACTTCTCTTCCGGAATATATGCTACTATATGAAGTGG 569
227 ThrTyrgValPheTyrgPhePhe 233
570 ACTTACCTCTCTATCTTTT 590
```

seq_name: gb_est2:BF124137

```
seq_documentation_block:
LOCUS      BF124137
DEFINITION 601760838F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4024064 5',
            mRNA sequence.
ACCESSION  BF124137
VERSION    BF124137.1 GI:10963177
KEYWORDS   EST.
SOURCE     Mus musculus
            house mouse.
ORGANISM   Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 904)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-re@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
```

```
Plate: LLAM9283 row: k column: 09
High quality sequence stop: 680.
FEATURES
    source
        1..904
            /organism="Mus musculus"
            /strain="C57/B6"
            /db_xref="taxon:10090"
            /clone="IMAGE:4024064"
            /clone_lib="NCI_CGAP_Mam5"
            /tissue_type="tumor, gross tissue"
            /dev_stage="7 months"
            /lab_host="DH10B"
            /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
            Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
            Library constructed by Life Technologies. Investigators
            providing samples: Lothar Hennighausen/Robin Humphreys,
            NIH"
BASE COUNT  193 a   214 c   276 g   221 t
ORIGIN
alignment_scores:
    Quality: 837.00      Length: 223
    Ratio: 4.024        Gaps: 0
    Percent Similarity: 93.274    Percent Identity: 82.063
alignment_block:
US-09-776-865-4 x BF124137
Align seg 1/1 to: BF124137 from: 1 to: 904
9 AlaProSerAspGlyGluGlySerAspArgThrProLeuLeuGlnAr 25
54 GCGGAAACGAGGATGAGAGAGCTCGACAGCACCCGCTCTCGCCGG 103
25 gAlaProArgAlaGluProAlaProValCysSerAlaArgTyrrAsnL 42
104 CGCCCGCAGACCCGAGCGGCTCCAGTGTGCTGCTGCTGCTGCTGCT 153
42 euAlaPheLeuSerPhePheGlyPhePheValLeuTyrgSerLeuArgVal 58
154 TAGCGAT.TTGGCGCTTCTGTGGTTTCTGCTGCTATGCTTACGGGTG 202
59 AsnLeuSerValAlaLeuValAspMetValAspSerAsnThrTrpAlaLy 75
203 AACCTGAGTGT.GGGTTAGTGGACATGCTAGATTCAATACAACTCTGAC 251
75 sAspAsnArgThrSerTyrgGluCysAlaGluHisSerAlaProIleLysV 92
252 TGATAATAGAACGCTCTAAGGAGTGTGCGGAACATCTGCCCCCATAAAG 301
92 allLeuHisAsnGlnThrGlyLysLysTyrgArgTrpAspAlaGluThrGln 108
302 TTCACCAATACACAGGTAAAGATACAGTGGGATGAGGAGAACTCAA 351
109 GlyTrpIleLeuGlySerPhePheTyrgTyrlleIleThrGlnIlePr 125
352 GGGTGGATTCTCGGCTCTTTTTCACGGCTACATCGTACCCAGATTCC 401
125 cGlyGlyTyrgValAlaSerArgSerGlyGlyLysLeuLeuGlyPheG 142
402 CGGTGGGTACATTGCCAGCAGGCTCGGAGGAGAGCTGCTGCTGGGCTGG 451
142 lyllePheAlaThrAlaIlePheThrLeuPheThrProLeuAlaAlaAsp 158
452 GCATCTTAGGACCTCCCTCTCTACCTGTTCACCCGCTGGCGCGAGAG 501
159 PheGlyValGlyAlaLeuValAlaLeuArgAlaLeuGluGlyLeuGlyGl 175
502 TTAGGCGGTGAGTCTCTGTTGCTTTAGAGCGCTGGAAGGAGTGGGAGA 551
175 uGlyValThrTyrgProAlaMetHisAlaMetTrpSerSerTrpAlaProp 192
```


142 lyllePhealathralallePheThrLeupPheThrProLeuAlaAlaAsp 158
 486 GCATCTTAGGCACCTCGTCCTACACCTGTCACACCGTGGCGGAC 535
 159 PheGlyValGlyAlaLeuValAlaLeuArgAlaLeuGluGlyGlyG 175
 536 TTAGGCGTGGTACCTCGCTGTGTAGAGCGCTGGAAGGACTGGGAGA 585
 175 uGlyValThrTyProAlaMetHisAlaMetTrpSerSerTrpAla.Pro 191
 586 GGGTGTAGCTTCCAGCTATGCACGCCATGCTGCTCTCCCTGGCTCCC 635
 192 ProLeuGluArg.SerLysLeuLeuSerLysSerTyAlaGlyAla 206
 636 CCTCTGGAAGAAGGCAAGCTTCTTACCATTTCCTATGCGGGGAGC 681

seq_name: gb_est2:BG541099

seq_documentation_block:
 LOCUS BG541099 893 bp mRNA linear EST 03-APR-2001
 DEFINITION 602570265F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4694680 5',
 mRNA sequence.

ACCESSION BG541099
 VERSION BG541099.1 GI:13533332
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 893)
 AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: L1CM1519 row: a column: 17
 High quality sequence stop: 549.
 Location/Qualifiers

FEATURES
 source
 1..893
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4694680"
 /clone_lib="NIH_MGC_77"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggcgctcgcc); Site_2: SfiI (ggccattatgcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence:
 5'-ATTCTAGAGCCGCGCGGCACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average inserted size 1.9
 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC Library."
 BASE COUNT 236 a 169 c 214 g 274 t
 ORIGIN

alignment_scores:
 Quality: 808.00 Length: 324
 Ratio: 3.848 Gaps: 8
 Percent Similarity: 64.815 Percent Identity: 54.321

alignment_block:
 US-09-776-865-4 x BG541099 ..

Align seg 1/1 to: BG541099 from: 1 to: 893
 162 GlyAlaLeuValAlaLeuArgAlaLeuGluGlyLeuGlyGlyValPh 178
 1 GGACCACTCATTTGTACTACAGCACTAGAGGACTAGGAGGGGTTCAC 50
 178 rTyProAlaMetHisAlaMetTrpSerSerTrpAlaProProLeuGlu 195
 51 AUTTCAGCCATGCATGCCATGCTGGTCTCTTGGGCTCCCTCTTGA 100
 195 rGSerLysLeuLeuSerLysSerTyAlaGlyAlaGlnLeuGlyThrVal 211
 101 GAAGCAAACTTCTTAGCATTCATATGCAGGACACAGCTGGGACAGTA 150
 212 ValSerLeuProLeuSerGlyValIleCysTyTrpMetAsnTrpThrTy 228
 151 ATTTCCTCTCTCTTCTGGAATAATTTGCTACTATATGAATGGACTTA 200
 228 rValPheTyPhePheGlyIleValGlyIleLeTrpPheIleLeuTrpI 245
 201 TCTCTTCTACTTTTGGTACTATTTGGAATATTG...GTTCTCTTTG 247
 245 leCysLeuValSerAspThrProGluThrHisLysThrIleThrProTy 261
 248 TCTGGTTAGTTAGTACACACCAACAAAAACACAGAGAAATTTCCCAT 297
 262 GluLysGluTyTrpIleLeuSerSerLeuLysAsnGlnLeuSerSerGln 278
 298 GAAAGGANTACATCTTTCATCATTAAGAAATCAG..... 333
 278 sSerValProTrpIleProMetLeuLysSerLeuProLeuTrpAlaIle 295
 333 333
 295 aValAlaHisPheSerTyAsnTrpThrPheTyThrLeuLeuThrLeu 311
 333 333
 312 LeuProThrTyMetLysGluValLeuArgPheAsnIleGlnGluAsnGl 328
 334AATGG 338
 328 yPheLeuSerAlaValProTyTrpLeuGlyCysTrpLeuCysMetIleLeu 345
 339 GTTTTATCTTCATTCCTTATTAGGCTCTTGGTATTATGATGATCCTCT 388
 345 exGly.GlnAlaAlaAspAsnLeuArgAlaArgTrpAsnPheSerThrle 361
 389 CTGGTCCAGCTGCTGCACAAATTTAAGGGCAAAATGGAATTTTCACTTT 438
 361 uTrpValArgArgValPheSerLeuIleGlyMetIleGlyProAlaIleP 378
 439 AVGTGTCGCAGAAATTTTAGCTTATAGGAATGATGGACCTGCAGTAT 488
 378 heLeuValAlaAlaGlyPheIleGlyCysAspTyTrpSerLeuAlaVal 394
 489 TCCTGGTAGCTGCTGGTTCATTTGGTGTGATTATTCTTTGGCCGCTGT 538
 395 PheLeuThrIleSerThrThrLeuGlyPheCysSerSerGlyPheSe 411
 539 TCTAATACTATATCAACAACT...GGAGGCTTTTGGTCTCTCTGGATTAG 585
 411 rIleAsnHisLeu.AspIleAlaProSerTyAlaGlyIleLeuLeuGly 427
 586 ATCCACCCATTTGGGATATG..... 605
 428 IleThrAsnThrPheAlaThrIleProGlyMetIleGlyProIleIleAl 444
 606GTGCTTCCACACACTGTTGGGAGAAATGGGAAC 637
 444 aArgSerLeuThrProGluAsnThrIleGlyGluTrpGlnThrValPheC 461
 638 CGGGTCTCTT.....ATGGTGGGGGAATTAAGTCT..... 668


```
/organism="Canis familiaris"
/db_xref="taxon:9615"
/clone_lib="Canine Brain cDNA Library"
/sex="Male"
/note="Organ: Brain; Vector: pSPORT1 (Gibco BRL); Site: 1:
Noti; Site: 2: SalI; Tissue was taken from the frontal,
occipital, temporal and parietal lobes, olfactory bulb,
hippocampus, cerebellum, thalamus, hypothalamus, midbrain
, pons, and medulla."
BASE COUNT 164 a 123 c 127 g 211 t 8 others
ORIGIN

alignment_scores:
  Quality: 796.00      Length: 222
  Ratio: 4.303        Gaps: 5
  Percent Similarity: 83.333  Percent Identity: 72.072

alignment_block:
US-09-776-865-4 x BI817031
Align seg 1/1 to: BI817031 from: 1 to: 633

237 GlylleIleTrpPheIleLeuTrpIleCysLeuValSerAspThrProG1 253
|||||
1 GGCAATCATTTGGTTATTTATTTATGGATCTTTTATAGTTAGTGAACCTCCAGA 50

253 uThrHisLysThrIleThrProTyrgLulysGluTyrlleLeuSerL 270
|||||
51 GACTCACAGAACATCTCCCGCGGAGAAAGAAATATATCTTTTCATCAT 100

270 euLysAsnGlnLeuSerSerGlnLysSerValProTirPileProMetLeu 286
|||||
101 TAAAAAATCAGCTTTCATCACAGAGTCAGTGCCTGATGATACCATGCTA 150

287 LysSerLeuProLeuTrpAlaIleValAlaHisPheSerTyAsnTr 303
|||||
151 AAATCACTGCGCTTTGGGCTATTGTAGTAGCATATTTTCTTCAACACTG 200

303 pThrPheTyTrThrLeuLeuThrLeuProThrTyTrMetLysGluValL 320
|||||
201 GACTTTTACACTTTATGACGTTATGCTTACATGAGGAGATCC 250

320 euArgPheAsnIleGlnGluAsnGlyPheLeuSerAlaValProTyTrLeu 336
|||||
251 TAAGGTTCAATGTTCAAGAGAATGGCTTCTATCTGCACCTGCCITACTT 300

337 GlyCysTrpLeuCysMetIleLeuSerGlyGlnAlaAlaAspAsnLeuAr 353
|||||
301 GGCTGTTGGTTATGATGATCTTCTGCTGCTCAAGCTGCTGACAAATTAAG 350

353 gAlaArgTrpAsnPheSerThrLeuTrpValArgValPheSerLeuI 370
|||||
351 GGCAAAATGGAATTTTCAACAATATGTGTCGAGAGAGATTTTATGCTTAA 400

370 leGlyMetIleGlyProAlaIlePheLeuValAlaAlaGlyPheIleGly 386
|||||
401 TAGGAATGATTGGACCTGCGGTATTCCTGGTAGCGCTGGATTATAGGT 450

387 CysAspTyTrSerLeuAlaValAlaPheLeuThrIleSerThrThrLeuG 403
|||||
451 TGTGACTATTTCATTTGGCGGTGGGT...CTTAACCATATCACACACTGG 497

403 lYGlyPheCysSerSerGlyPheSerIleAsnHisLeuAspIleAlaPro 419
|||||
498 GANGCTTTGCTCTTNT..... 514

420 SerTyralaGlyIleLeuLeuGlyIleThrAsnThrPheAlaThrIleP 436
|||||
515 .....GGATTAAACATCAANCATCTGGATGATGTC 543

436 roGlyMetIleGlyProIleIleAlaArgSerLeuThrProGluAsnThr 452
|||||
|||||
```

[illegible]